



Examining Early Child Development in Low-Income Countries:

A Toolkit for the Assessment of Children
in the First Five Years of Life

Lia C. H. Fernald, Ph.D.
Patricia Kariger, Ph.D.
Patrice Engle, Ph.D.
Abbie Raikes, Ph.D.

Examining Early Child Development in Low-Income Countries:

A Toolkit for the Assessment of Children
in the First Five Years of Life

Lia C. H. Fernald, Ph.D.

Patricia Kariger, Ph.D.

Patrice Engle, Ph.D.

Abbie Raikes, Ph.D.



THE WORLD BANK
Washington, D.C.

© 2009 The International Bank for Reconstruction and Development / The World Bank
1818 H Street NW
Washington DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org
E-mail: feedback@worldbank.org

All rights reserved

The findings, interpretations, and conclusions expressed herein are those of the author(s) and do not necessarily reflect the views of the Executive Directors of the International Bank for Reconstruction and Development / The World Bank or the governments they represent.

The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Rights and Permissions

The material in this publication is copyrighted. Copying and/or transmitting portions or all of this work without permission may be a violation of applicable law. The International Bank for Reconstruction and Development / The World Bank encourages dissemination of its work and will normally grant permission to reproduce portions of the work promptly.

For permission to photocopy or reprint any part of this work, please send a request with complete information to the Copyright Clearance Center Inc., 222 Rosewood Drive, Danvers, MA 01923, USA; telephone: 978-750-8400; fax: 978-750-4470; Internet: www.copyright.com.

All other queries on rights and licenses, including subsidiary rights, should be addressed to the Office of the Publisher, The World Bank, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2422; e-mail: pubrights@worldbank.org.

**Examining Early Child Development in Low-Income Countries:
A Toolkit for the Assessment of Children
in the First Five Years of Life**

Authors

Lia C. H. Fernald, Ph.D.

Patricia Kariger, Ph.D.

Patrice Engle, Ph.D.

Abbie Raikes, Ph.D.

Prepared for the World Bank Human Development Group, April 3, 2009

Corresponding author: Patricia Kariger (patriciakariger@gmail.com)

Acknowledgements: We are grateful for a critical review of this document by Frances Aboud (McGill University), Santiago Cueto (Catholic University, Peru), Ed Frongillo (University of South Carolina), Jane Kvalsvig (University of Kwa-Zulu Natal, South Africa), Ann Weber (University of California, Berkeley), Paul Wassenich (University of California, Berkeley), Michelle Neuman (The World Bank), and Mary Eming Young (The World Bank). Thanks to Robin Dean, Kallista Bley and Anna Moore for research assistance.

TABLE OF CONTENTS

TABLE OF CONTENTS.....	v
TABLE OF FIGURES.....	viii
EXECUTIVE SUMMARY	ix
CHAPTER 1: IMPORTANCE OF MEASURING CHILD DEVELOPMENT.....	1
Introduction.....	1
Conceptual framework.....	4
Development across childhood.....	5
Differential vulnerability	7
Cultural norms and development.....	9
Cumulative risk.....	11
Environmental context of development.....	13
CHAPTER 2: DOMAINS OF DEVELOPMENT TO BE MEASURED	15
Introduction.....	15
Cognitive skills	15
Executive function	17
Language skills	18
Motor skills.....	19
Social/emotional	20
CHAPTER 3: THEORETICAL DECISIONS IN SELECTING INSTRUMENTS.....	22
Introduction.....	22
Purpose of assessment.....	22
Types of assessments	24
Direct Tests.....	26
Ratings and Reports	28
Observational Measures.....	30
Screening tests versus assessment of abilities	31
Screening tests	31
Ability tests	33
Population versus individual-based testing.....	33
Rationale for population-based tests.....	34
EDI: example of a population-based assessment.....	34
Survey-based measures.....	35
Ethical risks and responsibilities in assessing young children.....	36
Constraints to consider.....	37
CHAPTER 4: MODIFICATION, ADAPTATION & STANDARDIZATION OF EXISTING TESTS.....	41
Introduction.....	41
Norming and milestones	41
Validity of measures	42
Lack of information on normative developmental milestones	42
Modification and adaptation	43
Preparatory work for test adaptation.....	44
Steps for successful test adaptation.....	45
CHAPTER 5: CREATION OF NEW TESTS	50

Introduction.....	50
Requirements for creating a new test.....	50
The “Standards” approach	53
Linking child-level Standards to program standards	56
Using Standards for assessment of learning	57
Pros of Standards approach.....	57
Cons of Standards approach.....	57
CHAPTER 6: TRAINING AND QUALITY CONTROL.....	59
Introduction.....	59
Connection with local psychologist.....	59
Inter-rater reliability	59
How to test inter-rater reliability.....	59
How to test rater accuracy.....	60
CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS.....	62
<i>Conclusions and future work</i>	62
<i>Broad recommendations</i>	63
<i>Specific recommendations</i>	67
Infants/Toddlers (Birth to 36 months)	68
<i>Primary recommendation for individual assessment</i>	68
<i>Alternative measures for individual assessment</i>	69
<i>Alternative measures for screening purposes</i>	70
<i>Summary: Recommended tests for children 0-2</i>	72
Preschool-aged Children (3 to 5 years).....	74
<i>Cognitive or comprehensive assessments</i>	74
<i>Language only</i>	78
<i>Motor skills</i>	79
<i>Executive function</i>	79
<i>Social and behavioral development</i>	80
<i>Summary: Recommended tests for pre-school children</i>	82
REFERENCES	86
APPENDIX A: Details of published and normed measures.....	103
Summary of options for published and normed measures.....	103
Achenbach Child Behavior Checklist (CBCL) 1 1/2-5 and Caregiver-Teacher Report Form (C-TRF).....	104
Ages and Stages Questionnaire (ASQ).....	105
Bayley Scales of Infant Development (BSID-I, 1st edition; BSID-II, 2nd edition; BSID-III, 3rd edition).....	106
British Ability Scales (BAS).....	107
Denver Developmental Materials II (formerly DDST)	108
Infant and Toddler Socio-Emotional Assessment (ITSEA, or BITSEA – brief form).....	109
Kaufman Assessment Battery for Children (Kaufman ABC).....	110
Leiter-R or Leiter International Performance Scale.....	111
MacArthur Child Development Inventory (CDI)	112
McCarthy Scales of Children’s Abilities (MSCA)	113
Pegboard	114
Preschool Language Scale (PLS-4)	115

Peabody Picture Vocabulary Test (PPVT)/Test de Imagenes Peabody (TVIP).....	116
Reynell Language Development Scales.....	117
Stanford Binet Intelligence Scale.....	118
Strengths and Difficulties Questionnaire (SDQ)	119
Wechsler Intelligence Scales for Children (WISC).....	120
Wechsler Preschool and Primary Scales of intelligence (WPPSI)	121
Woodcock-Johnson (or Woodcock-Muñoz) (WJ).....	122
APPENDIX B: New tests from developed countries	123
Australian Early Development Index (Australia)	123
Early Development Instrument (Canada)	123
APPENDIX C: New tests from developing countries	124
Cambodian Developmental Assessment Test (Cambodia).....	124
Early Childhood Care and Development Checklist (Philippines)	124
Escala Argentina de Inteligencia Sensorimotriz (EAIS)(Argentina).....	124
Escala de Evaluación del Desarrollo Psicomotor (EEDP) (Chile)	125
Grover-Counter Scale of Cognitive Development, Revised (South Africa).....	125
Guide for Monitoring Child Development (Turkey)	125
ICMR Psychosocial Developmental Screening Test (India)	126
IEA Preprimary Program Assessments (Multi-national).....	126
Kilifi Developmental Inventory (Kenya).....	127
Kilifi Developmental Checklist (Kenya)	127
Parental Report Scales (Tanzania, Nepal)	127
Shoklo Developmental Test (Thailand).....	128
Shoklo Neurological Test (Thailand).....	128
Test de Desarrollo Psicomotor (TEPSI) (Chile)	128
Yoruba Mental Subscale (Nigeria)	129
APPENDIX D: Details of tests developed to measure executive function	130
Bayley Examiner Assessment.....	130
Backward Digit Span Test	130
Behavioral Rating Inventory of Executive Function-P (BRIEF-P)	130
Delay of Gratification Test	131
Leiter Examiner Report.....	131
Stroop test (adapted for younger children as Day/Night test).....	131
Wisconsin card-sorting task.....	132
APPENDIX E: Table of measures used and where	133

TABLE OF FIGURES

Figure 1: Pathways connecting poverty and poor child development.	4
Figure 2: Timing of human brain development, from (Grantham-McGregor, et al., 2007).....	5
Figure 3: Adapted representation of Bronfenbrenner’s ecological model of child development (Wortham, 2007).....	12
Figure 4: Assessing inter-rater reliability	60
Figure 5: Testing accuracy.....	61
Figure 6: Flowchart for decision-making regarding assessment of early childhood development.	66

EXECUTIVE SUMMARY

The primary purpose of this toolkit is to provide a resource for researchers from various disciplines interested in planning and evaluating programs or interventions aimed at improving the health and development of infants and young children. The toolkit aims to: provide an overview of issues affecting early development and its measurement; discuss the types of tests typically used with children under five years; provide guidelines for selecting and adapting tests for use in developing countries, and make recommendations for planning successful assessment strategies. The toolkit focuses on children who have not yet entered school, and are thus under six years old.

Early childhood is characterized by developmental spurts and plateaus and tremendous bio-behavioral shifts. The various “domains” of development – cognitive, language, executive function/self-regulatory, motor, and social/emotional – all contribute to long-term well-being. It is thus essential to capture effects of early intervention programs on *multiple* domains of child development in order to capture a wide spectrum of abilities.

Deciding “why” to measure children’s development, “what” to measure and “how” to measure child development outcomes are crucial steps in the evaluation of interventions and programs targeting young children. In thinking about how to answer these questions, it is critical to examine: 1) the purpose of the testing; 2) the difference between screening and assessment of abilities and achievement; 3) the different modes of testing available, and 4) the use of population level vs. individual level testing. There may be practical and logistical issues that will affect the selection of tests. Intervention teams may find it necessary to consider which tests suit the project best and are feasible given constraints such as: budget, copyright issues, ethical issues, time allocated for testing, training, test setting, capacity of the respondent, language and cultural differences and materials.

While many language and cognitive-developmental assessments have been used in the developed world, most assessments must be modified substantially before use in developing countries. The steps for appropriate adaptation of tests are to produce an accurate translation of the test and the underlying construct(s), to adapt the test content to the local context, to adapt the test administration procedures to the local context, conduct pilot tests, and then undergo a

process of iterative adaptation and testing of the assessment. Alternatively, some research teams may decide to create their own tests rather than adapting existing tests. Successful generation of new tests involves an inter-disciplinary research team, an adequate representative sample for testing items and test cohesion, and the concurrent development of norms or standards that represent typical development. One approach to new test development is to generate tests based on locally agreed-upon “standards” for what children know and are able to do at a particular age (Early Learning and Development Standards).

Finally, we recommend several assessments because they measure a variety of domains; are psychometrically adequate, valid and reliable; have enough items at the lower end to avoid having some children fail all items; are enjoyable for children to take (e.g. interactive, colorful materials); relatively easy to adapt to various cultures; have been shown to discriminate developmental differences among children under study in various contexts; are easy to use in low-resource settings, e.g. not requiring much material; are not too difficult to obtain or too expensive; and are able to be used in a wide age range. Before using any test recommended here, however, we suggest that every test should be carefully examined and piloted within the given cultural context, and with the collaboration of a local psychologist, to be sure that it is measuring what it intends to measure.

CHAPTER 1: IMPORTANCE OF MEASURING CHILD DEVELOPMENT

Introduction

The psychological and biological changes that occur as a child transitions from a dependent infant to an autonomous teenager are collectively referred to as *child development*. These changes include the development of language, cognitive skills (e.g., symbolic thought, memory, and logic), social-emotional skills (e.g., a sense of self, empathy and how to interact with others) and motor skills (e.g., sitting, running, and more complex movements, etc.). It is now well accepted that development is a process that is not determined independently by nature or nurture alone, but by “nature *through* nurture” (pp. 41) (Shonkoff & Phillips, 2000). Changes throughout development result from multidirectional interactions between biological factors (genes, brain growth, neuromuscular maturation), and environmental influences (parent-child relationships, community characteristics, cultural norms) over time (Gottlieb, 1991; Pollitt, 2001; Shonkoff & Phillips, 2000). These interactions lead to the re-organization of various internal systems that allow for new developmental capacities (Thelen, 2000). For example, the emergence of locomotive skills results from the co-occurrence and interactions among physiological systems (muscle strength; the ability to balance), social-emotional change (the motivation to move independently), and experience (adequate opportunity to “practice” the emerging skill) (Adolph, 2002; Adolph, Vereijken, & Denny, 1998; Adolph, Vereijken, & Shrout, 2003). The conceptualization of development as a dynamic interplay between biological and environmental factors suggests that development is malleable and can be enhanced by interventions affecting the child, the environment or both.

The primary purpose of this toolkit is to provide a resource for researchers and program personnel from various disciplines interested in planning and evaluating interventions aimed at improving the development of infants and young children. The toolkit provides: an overview of issues affecting early development and its measurement; a discussion of the types of assessments typically used with children five years and under; guidelines for selecting and adapting tests for use in developing countries, and recommendations for planning successful assessment strategies. Our recommendations are primarily based on tests that have already been adapted for use in

other countries, in spite of the fact that many studies do not report on the adaptation process. Furthermore, we focused on tests that have been shown to discriminate successfully between groups of children (e.g. those who received a nutrition/health/early childhood intervention).

For the purposes of this review, the toolkit will emphasize the assessment of children aged five and under for several reasons. The primary reason we are focusing on this age group is that during the first five years of life, children's language, early understanding of mathematics and reading, and self-control emerge. The extent to which children master these skills during this critical period has implications for success in school (Lerner, 1998), and thus we wanted to focus on children in this pre-school period. Given that children in some lower- and middle-income countries enter school at later ages, however, the tests that are reviewed may also be appropriate for children who are slightly older (e.g. 6 or 7 years old).

The majority of the assessments reviewed and presented in this toolkit are for child-based measures that occur through an individual (one-on-one) assessment of a child. While we agree that assessments designed for the population-level are also necessary and important, there are few population-based measures of early childhood development that do not involve an individual assessment of a child. Thus, the majority of the recommendations presented in the toolkit can be adapted for use at the population level by examining the data in aggregate.

The toolkit is essential at this time for the following reasons:

- ❖ ***Children in developing countries are growing up at a disadvantage.*** The first paper in a recent Child Development series in the *Lancet* estimated that over 200 million children under 5 years worldwide are not fulfilling their potential for growth, cognition, or socio-emotional development (Grantham-McGregor, et al., 2007). During the first five years of life, children lay the groundwork for lifelong development (Shonkoff & Phillips, 2000). Thus, it is critical to assess children during this vulnerable period in order to determine if they are developing appropriately, and to develop interventions if the children are not developing optimally.
- ❖ ***Assessments of children must expand to include a wider range of outcomes.*** Interventions in early childhood are critically important not just for practical issues of feasibility and cost-effectiveness (Engle, et al., 2007), but also because of the brain's greater plasticity (i.e., capacity to benefit from environmental interventions) and physiologic development (Nelson, 2000). Most of the emphasis on evaluating outcomes in these circumstances is on height and

weight of children. Data gleaned from large studies including early child development assessments could be used to not only identify developmental differences, but also to advocate for a variety of services or interventions that would benefit child outcomes in the populations measured.

- ❖ *As far as we know, no such toolkit exists as present.* It is hoped that the information provided within this document will assist researchers and program personnel in the development and refinement of early childhood interventions. The accurate measurement of a young child's abilities – that may reflect future productivity – is essential for understanding the immediate and long-term impacts of such interventions.

Conceptual framework

Early childhood is the time of greatest risk and greatest opportunity.
Because young children have developing neuronal systems that are so plastic, children are simultaneously vulnerable to environmental influences and also capable of benefiting from interventions.

Poverty, socio-cultural factors, psychosocial and biological risk factors all work together to influence child development and long-term adult productivity (Grantham-McGregor, et al., 2007; Walker, et al., 2007) (**Figure 1**). A young child develops through advances in three inter-related domains: sensori-motor, socio-emotional, and cognitive and language abilities. The child's development is determined by the integrity and function of the central nervous system and by positive and negative environmental factors that affect development. Positive environmental factors include opportunities to explore surrounding and engage in learning activities; negative factors include exposure to psychosocial risks (e.g., harsh disciplinary techniques or maternal depression), and biological risks, such as malnutrition and infectious diseases.

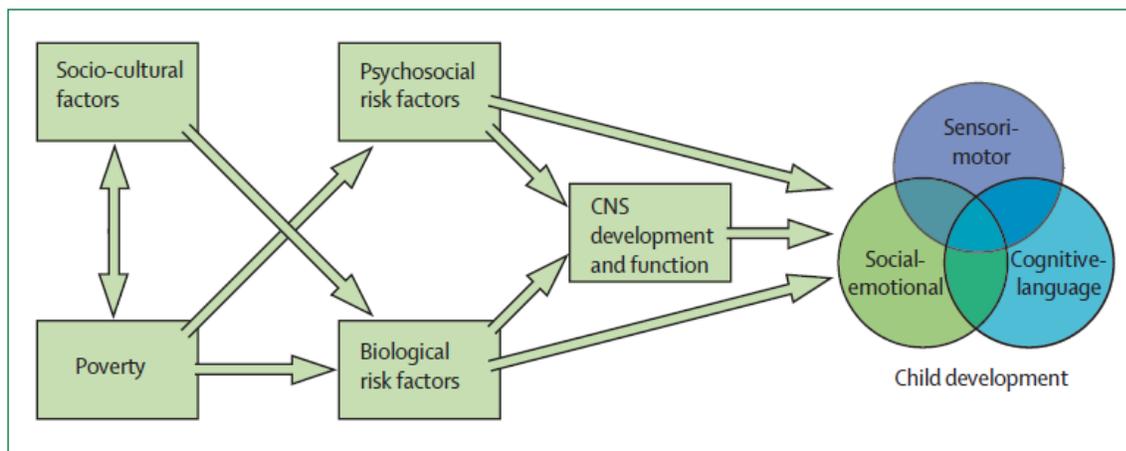


Figure 1: Pathways connecting poverty and poor child development.

From (Walker, et al., 2007)

Risk factors can influence development *directly* by affecting a child’s behavior -- for example, causing them to fuss more or play less -- and *indirectly*, by altering brain development and function. Poverty and socio-cultural factors, such as social marginalization, increase the likelihood of both types of risks. Preventing exposure to risks or intervening to reduce their effects on development enhances a child’s capacity to reach their developmental potential.

Development across childhood

Early childhood is characterized by developmental spurts and plateaus and tremendous bio-behavioral shifts (Bracken, 2007; Shonkoff & Marshall, 2000). Rapid brain and physical development, social relationships, and environments work together to create phenomenal advances in children’s abilities during this time frame (**Figure 2**). Children’s language, early understanding of mathematics and reading, and self-control emerge during the first five years, and the extent to which children master these skills has implications for success in school (Lerner, 1998).

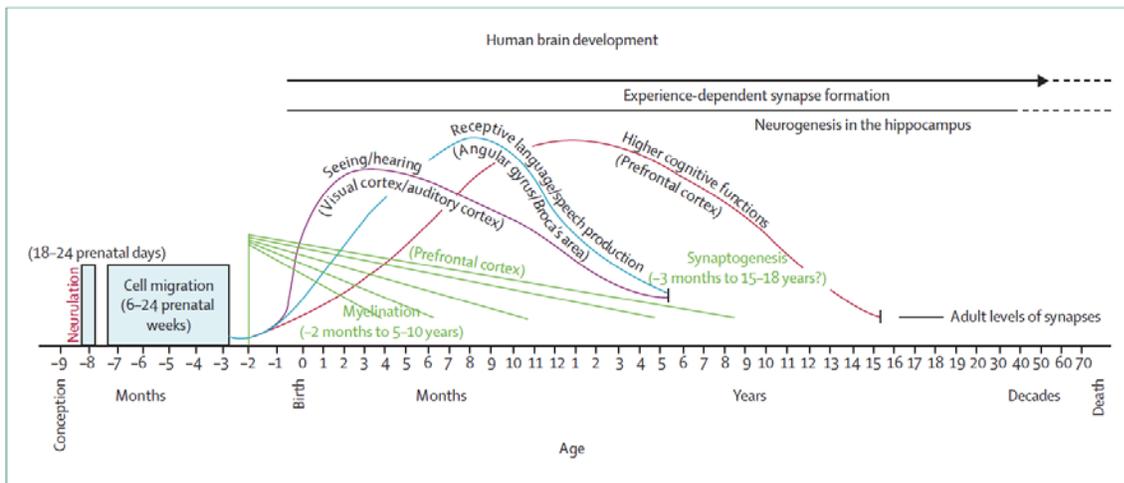


Figure 2: Timing of human brain development, from (Grantham-McGregor, et al., 2007)

New capacities emerge continually and often in close succession, as developments in one domain are catalysts for development in another (for example, after learning to walk, children are faced with new demands on self-control, as parents are more likely to restrict their behavior and expect that “no” will be obeyed). Similarly, children who are slow to develop in one domain (e.g. understanding language) may have limited capacity to display the skills that they possess in

other domains (e.g. cognitive tasks that require language skills). Thus, development in young children must be assessed as *comprehensively* as possible (Miesels & Atkins-Burnett, 2000).

Heckman has argued that non-cognitive skills as well as cognitive skills play a significant role in school achievement, productivity, and likelihood of becoming a criminal (Heckman & Masterov, 2005). Interventions to improve both non-cognitive and cognitive skills are most likely to have a long-term effect when intervention occurs early; the later the intervention, the more expensive to create a positive benefit. Therefore he argues strongly that preschool combined with home visiting and parenting support for disadvantaged children will have greater long-term economic benefits than other intervention programs (Heckman & Masterov, 2005).

Skills emerge at different rates and ages, especially in the first two years, and windows for “normative” attainment tend to be wide (e.g., 8-18 months of age for walking (WHO, 2006a, 2006b)). Within-individual variability is also common, and a child’s progression in any particular domain may be unstable or “bounce” around rather than advance steadily over time (Darrach, Redfern, Maguire, Beaulne, & Watt, 1998; Pollitt & Triana, 1999). For example, a child who crawls early may not necessarily walk early and vice versa. This variability reflects the fact that development results from interactions among child characteristics, environmental factors, the demands of the developmental task(s) at hand, and that during periods of rapid change, development tends to occur in one domain at a time.

The breadth and depth of behaviors that can be assessed increases with age, and the advancement in communication and other skills during the preschool years provides additional modes for testing (Snow & Van Hemel, 2008). Aptitudes important for cognition and school success – e.g. pre-literacy skills, attention and focus, memory, getting along with other children – can be measured at this age level. Children’s environments become more increasingly differentiated, and individual differences in abilities become more pronounced (Rydz, Shevell, Majnemer, & Oskoui, 2005; Shonkoff & Marshall, 2000). There is evidence that by the age of three or four years, preschoolers’ developmental test scores are strongly predictive of later performance on school achievement and intelligence tests (Neisser, et al., 1996).

When the emergence of a child’s ability is significantly slower than average for age, the child is considered to be “delayed” on that ability. At an early age, the rate of emergence of abilities differs considerably among children, and the delay may disappear. The ability often

emerges at an acceptable level later for some children, but for others the ability may continue to develop more slowly than age mates.

“Delay” is always determined relative to normative development within a given population; therefore, cut-off scores that define delay in one population cannot be assumed to define delay in another.

Delays as well as abilities become evident with age, and problems in specific areas are not apparent until the child reaches an age when those skills are typically learned and can be effectively evaluated (Glascoe, 2001; Rydz, et al., 2005). Thus, a child with no apparent delays in communication or cognitive skills at three years may nevertheless be diagnosed with reading difficulties at six years (Glascoe, 2001). Continued testing and tracking through school-age years is important for evaluating the long-term benefits of programs and interventions that begin early in life (Glascoe, 2001; Rydz, et al., 2005; Shonkoff & Marshall, 2000; Snow & Van Hemel, 2008).

Differential vulnerability

Infants (defined as children from 0-12 months old) and children one to three years old growing up in poverty are exposed to poor sanitation, crowded living conditions, lack of psychosocial stimulation and fewer household resources (Walker, et al., 2007). Young children growing up in poverty are more likely to experience developmental delays and growth deficits than those from more privileged backgrounds because they are disproportionately exposed to a wide range of co-occurring risk factors that impact development (Bolig, Borkowski, & Brandenberger, 1999; Bradley & Corwyn, 2002; Brooks-Gunn, Klebanov, Liaw, & Duncan, 1995). There is also increasing evidence that certain sections of the brain – such as those associated with language, memory and executive function – are affected by psychosocial and biological risk factors associated with poverty in US children (Noble, Tottenham, & Casey, 2005).

Children from low-income backgrounds in the United States are more likely to experience poor nutrition or malnutrition (Brooks-Gunn, et al., 1995; Guthrie, 1999), less stimulating learning environments (Bradley, Corwyn, McAdoo, & Garcia Coll, 2001; Brooks-

Gunn, Leventhal, & Duncan, 2000), more limited linguistic role models (Hart & Risley, 1995; Hoff, 2003), crowded or substandard housing (Evans & English, 2002; Koch, Lewis, & Quinones, 1998), exposure to domestic or community violence (Brooks-Gunn, et al., 2000; Hsieh & Pugh, 1999), and greater environmental hazards (Cohen, et al., 2003; Jacobs, et al., 2002; Karns, 2001; Wamboldt, et al., 2002). In the developing world the conditions contributing to poor development are exacerbated by more extreme poverty, poor sanitation, crowding, and even more limited access to resources (Guo & Harris, 2000). Not surprisingly, significant associations exist between low height-for-age (stunting) and delayed cognitive development (Agarwal, Upadhyay, Tripathi, & Agarwal, 1987; Bogin & MacVean, 1983; Chun, 1971; Clarke, Grantham-McGregor, & Powell, 1991; Cravioto, DeLicardie, & Birch, 1966; Florenco, 1988; Freeman, Klein, Townsend, & Lechtig, 1980; Jamison, 1986; Monckeberg, 1972; Mook & Leslie, 1986; Paine, Dorea, Pasquali, & Monteiro, 1992; Powell & Grantham-McGregor, 1980; Sigman, Neumann, Jansen, & Bwibo, 1989), psychomotor development, (Lasky, et al., 1981; Monckeberg, 1972; Powell & Grantham McGregor, 1985; Sigman, et al., 1989) poor fine motor skills, (Cravioto, et al., 1966; Grantham McGregor, Walker, Chang, & Powell, 1997) and altered behavior (Fernald & Grantham-McGregor, 1998).

Children in the developing world are more likely to be vulnerable to deficiencies in basic health and nutrition than are children in the developed world, and these deficiencies contribute to delayed physical and cognitive development.

Research on low-income children in the US suggests that developmental scores are in the normal range during infancy, but then the scores decline during the preschool years; this pattern is not apparent in middle-income samples (Black, Hess, & Berenson-Howard, 2000). Similar findings are reported in Ecuador (Schady, 2006), Jamaica (Powell & Grantham McGregor, 1985), and Ethiopia (Aboud & Alemu, 1995). These data suggest that low income children are increasingly vulnerable to the external environment as they enter their second and third year of life (Gottlieb, 1991; Werner, 2000). Differences begin to emerge as children learn more complex processes, such as language. This means that a less stimulating environment can support development in the first six months but not as the children continue to develop. As they grow up, children living in poverty in the developing world are likely to have substantially lower

wages than healthier adults (Boissiere, Knight, & Sabot, 1985), and are thus less likely to be able to provide increased stimulation and resources for their own children, thereby perpetuating the cycle of poverty (Sen, 1999).

Cultural norms and development

Culture refers to a set of beliefs, values, goals, attitudes and activities that guides the manner in which a group of people live (Payne & Taylor, 2002). Any particular culture is shaped by a broad spectrum of factors, such as geography, religion, political and economic structures, access to educational and health care systems, and the degree to which modern technology is present. Parenting practices and ideas about child development are largely determined by cultural ideals. Cross-cultural studies of development aim to distinguish which skills and abilities are universal from those that are culture-specific or are unique to an individual (Carter, et al., 2005).

Cultures have a wide range of values for the skills and abilities that children should develop and when they should be exhibited (i.e., “norms” or normative ages when skills are typically displayed). Abilities may emerge earlier if they are valued and encouraged in a particular culture. However, this does not mean that the ability will not emerge at some point. These culturally specific patterns must be considered in assessing the validity of a measurement, and are of particular concern when comparisons are made across population/ethnic groups or across countries. When comparisons are made within a group (e.g., intervened vs. control) the concern is limited to being sure that the assessment used is actually measuring the capacities that the intervention was designed to change.

These concerns are greater if one is assessing intelligence (i.e., using a standardized IQ test) than if one is assessing the attainment of specific skills or abilities that are typically measured in young children. In almost all cultures, the kinds of skills that young children need for school do not vary. As school becomes more universal, the necessary skills become more consistent across cultures. These include not only academically-related skills, such as language and symbol recognition, but also social skills such as knowing how to function in groups, wait for a turn, or inhibit an initial response. These skills are useful not only for school but also for overall productivity and adaptability throughout later life.

There is no simple way to ensure cross-cultural comparability of early cognitive tests. An extreme position that suggests that each culture is totally unique and requires special assessment methods ignores the reality of *universal rights* such as the right to education and the right to survival and development for every child, as guaranteed in the Convention for the Rights of the Child. On the other hand a position that all children must be judged by exactly the same measurement – even if well adapted – ignores the wide range of *different values for and ways of learning* that results in some abilities developing more quickly in some cultures than others (e.g., using rules of social conduct and respect).

Some evaluations have attempted to relate scores on measures assessing skills necessary for children to do well in school and be productive as adults (e.g., literacy and problem-solving skills) with culturally valued attributes deemed important for being successful within a particular society (e.g., responsibility for carrying out tasks necessary for daily living). Among the Yoruba in Nigeria, very young children who were rated as more responsible by parents to purchase items or retrieve particular objects scored higher on a modified (shortened) version of the Bayley Scales of Mental Development (Bayley, 1969) than did children with lower responsibility ratings (Ogunnaike & Houser Jr., 2002). This suggests that the two types of measures were related.

In Zambia, adult ratings of a school-age child's capacity to complete specific tasks were highly predictive of school grade completion and adult literacy scores; however, this finding was true only for girls (Serpell & Jere-Folotiya, 2008). The authors speculated that because girls participate more in domestic chores than do boys, adults may have had greater opportunity to observe and evaluate their abilities. In contrast, scores on other locally developed tests more rooted in Western notions of abilities were strongly predictive of grade attainment and literacy for boys (especially those living in urban areas), but these were not predictive for girls' later literacy scores (rural girls, in particular). Gender, demographic characteristics (rural vs. urban) and differences in schooling greatly influenced the findings. In rural Guatemala, test performance was associated with child behaviors -- in particular their ability to complete a series of three chores without additional instruction -- as well as with adults' ratings of children's "smartness" (Nerlove, 1974). These examples illustrate both the links between tests and local conceptions of ability, and the complexities of using local notions of attributes to predict later capacities, and highlight the need for scrutinizing all types of assessments.

Whether adapting existing tests or developing local measures, every effort must be made to ensure that tests are fair for all children assessed. Test fairness relates to the degree to which a measure is equally valid for individuals with different demographic characteristics, including access to resources and educational services, gender, culture, and ethnicity. Issues to consider include familiarity with the type of materials (writing, numbers, pictures), with the cultural relevance of items (e.g., horses are unfamiliar in Africa), testing situation (e.g., talking to an adult), and the importance of responding quickly. For example, Zambian children have extensive experience making objects from wire, but little experience with drawing. School-aged children asked to reproduce a wire model of an object (the Panga Munthu Test, based on the Goodenough-Harris Draw-a-Person test (Harris, 1963)) did so more effectively than when asked to draw a pictorial figure using paper and pencil, illustrating that the use of a familiar medium (i.e., wire vs. pencil and paper) was an important factor in the assessment of this skill (Ezeilo, 1978; Kathuria & Serpell, 1998).

While there are methodologies for adapting test items, materials and administrative procedures to make them as fair as possible, cross-cultural researchers acknowledge that the development of culture-free cognitive tests is impossible, as all tests (even non-verbal) are inherently biased, and all must be adapted (Cole, 1999; Greenfield, 1997; Rosselli & Ardila, 2003). Adaptations of assessments can at best produce a reduction in cultural differences in performance on any test (Anastasi & Urbina, 1997). Within these constraints, it is recommended to consider assessments that have successfully discriminated amongst groups of children in various cultural contexts, and to always bear in mind the necessity for careful selection and adaptation or development of assessments to evaluate young children.

Cumulative risk

Developmental outcomes are influenced by the number of biological, social and family risk factors impacting a child's development rather than the specific type or weighting of each factor (Breitmayer & Ramey, 1986; Rutter, 1979; Sameroff, Seifer, Baldwin, & Baldwin, 1993). Examples of risks might include poor infant nutrition, stressful life events, poor mother-child interactions, absence of father or other social supports, exposure to environmental risks, or

changes in family employment status; risks occur across the various domains of influence (Figure 3).

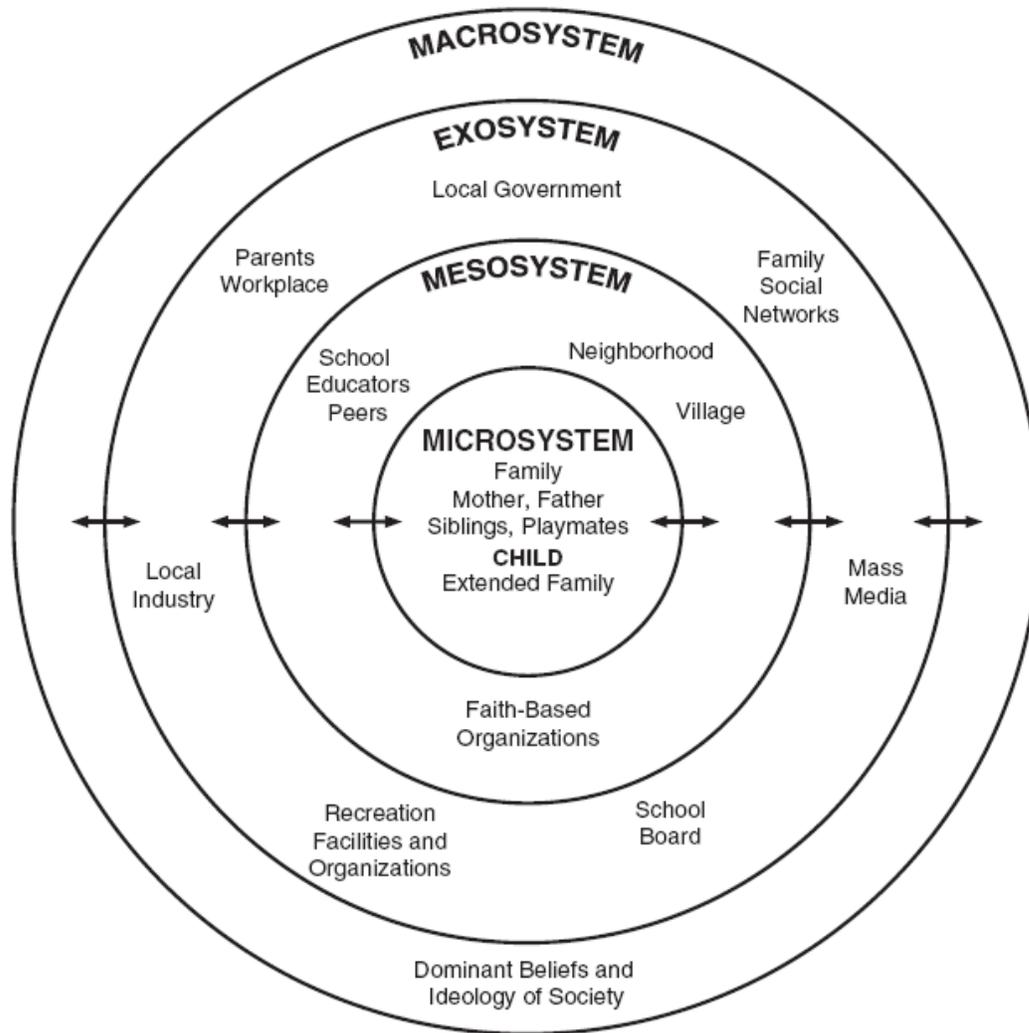


Figure 3: Adapted representation of Bronfenbrenner’s ecological model of child development (Wortham, 2007).

Children living in poverty are exposed to an increasing number of risks over time, and the cumulative effects of these risk factors on development become more evident as children get older. Prior work has found that higher cumulative levels of risk are related to poorer cognitive development (Brooks-Gunn, et al., 1995; Sameroff, et al., 1993; Sameroff, Seifer, Barocas, Zax, & Greenspan, 1987), psychological distress and behavior problems (Brooks-Gunn, et al., 1995; Evans, 2003), and communicative development and symbolic behavior (Hooper, Burchinal, Roberts, Zeisel, & Neebe, 1998).

For families in persistent poverty, many risks are present throughout the child's life and other risk factors may emerge and accumulate over time.

The addition of these kinds of risks to existing factors (e.g., low parent IQ and/or education, high family density, low birth weight children, parental mental health, or understimulating home environment) contribute to higher cumulative risk levels in children one to three years old than in young infants. Research demonstrating that children living in poverty are at risk for adverse developmental outcomes is likely due to the cumulative effects of exposure to risk factors rather than the result of any single explanatory mechanism.

There is good evidence that integrated interventions addressing multiple risks to children's development (e.g., health nutrition and stimulation) are more effective at preventing developmental decline than singular interventions in the developing world (Engle, et al., 2007). Yet, for practical reasons, interventions and programs often cannot address all poverty-related risks, and instead must prioritize activities that will have the largest impact on development in the population under study. Decisions about the best ways to intervene should be guided by the types of risks present; the percentage of children affected; the severity of the risks and research on the age at which children are most likely to benefit from interventions. Evaluations of programs and interventions must measure all existing risks and consider analytic strategies that will be most effective at demonstrating the desired impact.

Environmental context of development

There is substantial diversity in the types of achievements that children demonstrate during the first five years. Some developmental achievements are more "canalized" than others, meaning that they are on a particular trajectory in which both the nature and timing are strongly affected by biological maturation (Bretherton, Bates, Benigni, Camaioni, & Volterra, 1979; McCall, 1981). Walking and talking are examples of traits that all healthy individuals ultimately demonstrate in the early years, although the timing in which they emerge can vary according to environmental factors. As children grow older, meaningful individual differences emerge. Healthy development depends more and more on the quality of the environment children are in (Shonkoff & Phillips,

2000). Thus, it is reasonable to expect that children in impoverished environments will appear increasingly dissimilar from their higher socioeconomic peers as they grow older (Wagstaff, Bustreo, Bryce, Claeson, & WHO, 2004).

When children are not in stimulating and responsive environments, it is unlikely that they will demonstrate the same competencies as children in stimulating, rewarding environments.

CHAPTER 2: DOMAINS OF DEVELOPMENT TO BE MEASURED

Introduction

The various “domains” of development – cognitive, language, executive function/self-regulatory, motor, and social/emotional – all contribute to long-term well-being (Kuhn & Siegler, 1998). But while developmental tasks such as walking and learning letters may be divided into domains for categorical purposes, they are overlapping and mutually influencing in children. It is thus essential to capture effects of early intervention programs on *multiple* domains of children in order to capture a wide spectrum of abilities.

Developmental assessment scores obtained during infancy and the first three years of life do not always predict development in later years (Bracken, 2007; Snow & Van Hemel, 2008). Thus, much of the literature supports the use of comprehensive assessment in under twos for measuring concurrent abilities and identifying delay, but cautions using such scores for predicting future development (Bradley-Johnson & Johnson, 2007; Snow & Van Hemel, 2008). Some cognitive processes – working memory, inhibition and attention (see *Execution function* below) -- measurable as early as 2.5 years of age show moderate-strong correlations with intelligence scores (Bradley-Johnson & Johnson, 2007; Neisser, et al., 1996) and achievement during childhood and adolescence (Duncan, et al., 2007). While these types of measures have not been rigorously tested or standardized with large groups of children (Rydz, et al., 2005), they may provide meaningful complementary information to the comprehensive developmental assessments typically conducted with young children.

Cognitive skills

Cognitive skills encompass analytical skills, mental problem-solving, memory, and early mathematical abilities (M. H. Johnson, 1998). For infants and toddlers, early cognitive development involves problem-solving with objects, such as learning to stack or nest objects, and early understanding of math, demonstrated by such behaviors as sorting objects and knowing what it means when someone asks for “one” or “two” of something (Kuhn & Siegler, 1998). By

age 3, most children are capable of solving simple puzzles, matching colors and shapes, and also show awareness of concepts such as “more” and “less.” As children approach school-age, cognitive development broadens in scope and includes children’s early knowledge of numbers, including adding and subtracting, and their familiarity with letters and print (see *Language skills* below) (Schneider & Bjorklund, 1998). Indicators of cognitive development as children near school entry include knowledge of letters and numbers, ability to retain information in short-term memory, and knowledge of key personal information like one’s name and address. Standardized tests of reasoning, problem-solving, memory and mathematical abilities at the start of school are strong and reliable indicators of children’s cognitive development and are predictive of scores throughout childhood.

Research increasingly demonstrates that cognitive abilities may be as strongly affected by the quality of the environment as they are by genetics (Shonkoff & Phillips, 2000).

Children’s cognitive development in the first five years is dependent on the quality of their early environments and their relationships with caregivers. Children with responsive caregivers, and those who are in more stimulating environments, are more cognitively advanced at the start of school than children in less stimulating homes; parents who interact frequently with their children promote their cognitive, social and emotional development (Shonkoff & Phillips, 2000).

Genetic influences are generally considered to account for approximately half of the variance in cognitive abilities (Kovas, Haworth, Dale, & Plomin, 2007) based on studies of identical twins. Kovas et al. (2007) in the UK show that both high ability and learning disabilities appear to be variations on this pattern, not significantly different kinds of abilities. Genetic influences tend to be more consistent across ages than environmental influences. Although genetics play a role in a child’s developing abilities, evidence shows the importance of genetic/environment interactions in how those genes are expressed, and in the important role that environmental variations play in development and education. It is possible that these environmental influences are even more important in conditions of poverty, malnutrition, and ill health.

Executive function

The concept of executive function is relatively new -- from the past 20-30 years -- and results from neuropsychological research on the effects of damage to the frontal lobes (Jurado & Rosselli, 2007). While the field is still evolving, and definitions of executive function are variable, there is general agreement that executive function comprises *fluid* abilities or processes that are engaged when a person is confronted with a novel situation, problem or stimulus. These fluid abilities are distinct from *crystallized* cognition or knowledge of information (such as vocabulary) (Jurado & Rosselli, 2007). Executive function processes are believed to include impulse control, ability to initiate action, ability to sustain attention, and persistence.

Executive function is often classified as a subcategory of cognitive skills, in spite of the fact that both cognitive and emotional processes are involved. The more cognitive executive function processes are linked to dorsolateral regions prefrontal cortex and have been called “cool” processes – such as remembering arbitrary rules, and other non-emotional aspects of the task. “Hot” executive function processes have been linked to the ventral and medial regions of the prefrontal cortex and describe the more emotional aspects of executive function – those involving inhibition, or delaying gratification (Hongwanishkul, Happaney, Lee, & Zelazo, 2005). Thus, executive function processes straddle both the cognitive and social-emotional domains.

While the roots of children’s executive functioning are apparent in infancy, executive function develops considerably in early childhood, as the frontal lobe develops (Anderson, 1998). In young children (2+ years), some of the processes most commonly cited as measurable are working memory (e.g., holding information in mind for a short time, such as a series of numbers); inhibition of behavior or responses as demanded by the situation or task (e.g., not opening a box until a bell rings or inhibiting a response that was previously correct but no longer is) and sustaining attention as required or being able to switch attention as necessary (e.g., shifting focus from the color of a test stimulus to the shape of the stimulus) (Carlson, 2005).

Engagement of executive function skills enable humans to adapt to ever-changing contexts and are indispensable for success in school, work and day-to-day living (Hongwanishkul, et al., 2005). Recent research suggests that in the US, attention processes in the preschool years is associated with academic achievement (Duncan, et al., 2007). The negative effects of socio-economic status on children’s school readiness in the US are believed to

be mediated by attention processes, suggesting that low-quality environments affect cognitive development in part by decreasing children's abilities to attend (NICHD, 2003). Executive functioning components can be measured separately, but often it's the capacity to integrate or coordinate them *to solve a problem or reach a goal* that is most significant to assess (Welsh, Friedman, & Spieker, 2006). Tasks requiring the engagement of multiple processes are considerably more difficult than using only one process (Carlson, 2005) but are more likely to reflect real-life demands.

Language skills

Children's language development begins long before the emergence of the first word (Bloom, 1998). Early indicators of language development include babbling, pointing, and gesturing in infancy, the emergence of first words and sentences in the first two years, leading to an explosion of words between ages 2 and 3 years (Woodward & Markman, 1998). As children move into the preschool years, indicators of language development include children's production and understanding of words, their abilities to tell stories, identify letters, and their comfort and familiarity with books. Standardized assessments of children's vocabularies and their knowledge of letters and print at the start of school predict their reading scores throughout childhood. In cultures with a history of literacy, children who do well on language tests are those who know a number of literary words. However, in cultures that do not have a long history of literacy, there are other criteria that can be used. In some African cultures, for example, grammatically correct and creative use of alliteration and metaphor are the mark of a child who is linguistically advanced (Harkness & Super, 1977; Kvalsvig, personal communication).

Children's language skills are also critical for their success in school. Not only does reading build upon children's early vocabulary, children also must understand directions from teachers and be able to communicate their feelings and thoughts to others. Like cognitive and social/emotional development, language development is dependent on stimulating home environments and relationships. Low-income children in the United States build their vocabularies more slowly than higher income children and speak many fewer words than their higher-income counterparts by kindergarten (Hart & Risley, 1995). This pattern occurs in part because they receive less infant directed speech and also because the speech that they hear has

reduced lexical richness and sentence complexity, both of which contribute to vocabulary growth (Hart & Risley, 1992; Hoff, 2003). In addition, within low-income homes, adult speech is less responsive to children's signals, less directed to infants and used less in the course of shared attention and shared communication (Tamis-LeMonda, Bornstein, & Baumwell, 2001). Reading to children early in life also supports language development. Because children's language development is heavily dependent on their exposure to words and books in the home, children whose parents are not literate may develop speech and vocabulary more slowly (Fernald, et al., 2006).

Motor skills

Large (or gross) motor development refers to the acquisition of movements that promote an individual's mobility. While the age and sequence of motor milestone attainment may vary both within and across samples of children, nearly all healthy children will eventually acquire the capacity to walk and more advanced behaviors (e.g., running, jumping, etc.). Advancement in motor skills was once thought to be determined by brain and neuromuscular maturation alone (Gessell, 1946), but recent research indicates that other factors – such as physical growth, caregiving practices (i.e., swaddling or carrying) and the opportunities to practice emerging skills – also contribute to motor progression (Adolph, et al., 1998; Adolph, et al., 2003; Kariger, et al., 2005; Kuklina, Ramakrishnan, Stein, Barnhart, & Martorell, 2004).

For infants and young children, large motor skills include learning to walk and run, and for preschool-aged children, large motor skills include walking on a line, controlling movements in games, and jumping. Although the timing of most large motor skills is not indicative of future development, a failure to demonstrate these skills may indicate the presence of a developmental delay. For example, a child who does not walk at age two may have a developmental disorder that should be addressed, and tests of gross motor skills are created to identify children whose development is far behind expectations.

Fine motor skills, such as drawing and writing letters, involve eye-hand coordination and muscle control. The acquisition of fine motor skills is significant because through them children gain a new way of exploring the environment and thus fine motor skills contribute to developmental achievements (Bushnell & Boudreau, 1993). Fine motor skills include such

abilities as picking up objects and holding eating utensils. For preschool-aged children, fine motor includes the ability to hold a pencil, write and draw. Difficulties in motor skills can indicate the presence of neurological or perceptual problems.

Social/emotional

Social and emotional development has implications for many domains of children's development (Saarni, Mumme, & Campos, 1998). In the first two years of life, much of children's social and emotional development centers on relationships with caregivers. During these years, children learn whether they will be responded to by others and how much they can trust those around them. Learning to explore is a fundamental task of infants and toddlers, and they are more confident in their explorations when they are confident that their caregivers will be available when they return from their explorations. In the first two years, children also acquire early strategies for dealing with their negative feelings. Warm, responsive relationships with caregivers are essential for teaching children to trust, and for helping learn to deal effectively with frustration, fear and other negative emotions (Thompson & Raikes, 2006). Healthy infants and toddlers will show preferential attachments to caregivers, are eager to explore novel objects and spaces, and enjoy initiating and responding to social interactions.

In the preschool years, social and emotional development expands to include children's social competence (how well children get along with others, including teachers and peers), behavior management (following directions and cooperating with requests), social perception (how well children can identify thoughts and feelings in themselves and others), and self-regulatory abilities (emotional and behavioral control, especially in stressful situations). All of these skills are critical for children's success in school (Thompson & Raikes, 2006). Children who are not able to discern the thoughts and feelings of others are more likely to behave aggressively and experience peer rejection (Denham, et al., 2003), and children with both "internalizing" behavior problems characterized by depressed, withdrawn behavior, and "externalizing" behavior problems or aggressive, angry behavior are more likely to have difficulty in school (Rimm-Kaufman, Pianta, & Cox, 2000).

Indices of children's behavior problems have often been used in studies in the developing world. It is both quick and cost-effective to ask parents to respond to questionnaires regarding

their children's behavior problems, keeping in mind that reports of behavioral and socio-emotional problems are likely to be influenced by cultural norms. Results should be interpreted with caution if they indicate multiple behavior problems; the prevalence of both internalizing and externalizing behavior problems is quite low in most contexts. Measures of behavior problems alone generally yield few insights into children's social and emotional well-being, although these measures can be useful in cases of extreme psychological distress results (Atwine, Cantor-Graae, & Bajunirwe, 2005 ; Mulatu, 1995). Moreover, the absence of behavior problems should not be taken as an indication of social and emotional well-being. Instead, it is important to use measures that index children's social competencies as well as their problematic behavior.

CHAPTER 3: THEORETICAL DECISIONS IN SELECTING INSTRUMENTS

Introduction

Deciding “why” to measure children’s development, “what” to measure and “how” to measure child development outcomes are crucial steps in the evaluation of interventions and programs targeting young children. These questions are even more critical when there is not a local literature to guide decision-making. The remainder of this section will outline the major issues involved with selecting assessment instruments. These include 1) the purpose of the testing; 2) the difference between screening and assessment of abilities and achievement; 3) the different modes of testing available, and 4) the use of population level vs. individual level testing.

Purpose of assessment

The first step in selecting measures is to clarify the purpose for the assessment. Assessments of child development can be conducted for various reasons: to plan interventions or services; to monitor or evaluate the impact of early childhood education programs; to investigate the effect of interventions or programs on specific outcomes of interest; to design a curriculum for a particular child; or to diagnose and assess child progress. The rationale for testing should clearly link to objectives or goals that in turn will help guide which domains to measure, the types of tests and testing modes to use, and approaches for interpreting and using the test information (Snow & Van Hemel, 2008).

For example, consider a project concerned with examining the impact of an early parent-stimulation program on child development. The general goal of the project would be to determine whether children 6-24 months of age receiving the intervention perform better on developmental tests than children in a control group. In order to select the instruments that will best serve the purpose of the assessment, it would be essential to answer the following questions:

❖ *What dimensions of a child’s development do you expect to be affected by the intervention?*

For example, in the case above, the authors may hypothesize that the major impact of the

intervention will be on changes in the interactions between caregivers and children (e.g., increasing adult-child engagement in learning activities), that would subsequently benefit child performance on language, social-emotional and problem-solving tasks. It is important to consider measuring aspects of development that link to immediate as well as longer-term outcomes (e.g., grade completion or achievement scores, literacy, etc.).

- ❖ ***What are the mechanisms at work?*** The answer to this question will help with the question above. Through which (biological and/or environmental) mechanism(s) is the intervention expected to operate? What is already known about the functional mechanism linking stimulation, for example, with child performance on various aspects of development that could guide the choice of outcomes? Which processes are most influenced by the intervention and which biological or environmental risk factors present in the population under study need to be considered in planning and evaluating the intervention? How do these factors change with age (e.g., stimulation programs are generally more effective if started when children are very young)?
- ❖ ***What are key elements of the context that must be considered in selecting a test?*** These may include: urban or rural setting; level of poverty; parent education and literacy; language spoken in the home; risk factors to which children are likely exposed, and access and familiarity with the media required for the assessment (e.g., pencil and paper).
- ❖ ***At what level will the effect be measured?*** Are the evaluators most interested in demonstrating impact at the individual, household, community or population level?
- ❖ ***How will the sample be selected?*** Given the study design, is it necessary to test all children or will it suffice (and also be more feasible) to measure a sub-sample of the population? What sample size will be needed to provide sufficient statistical power to detect the anticipated effect, or to detect the minimum meaningful effect?
- ❖ ***What are the goals of the assessment/ evaluation?*** Is there an interest in showing relative improvement in one group over another or in individual improvement in developmental scores or in domains? Which measures have been shown in the literature to be most sensitive to detecting treatment effects in similar samples of children? Do these change with age?

❖ *What is your plan for analysis?* Are the assessments occurring in a context where norms (i.e., age-related references for the development of skills or abilities) are available? If so, are the norms relevant and/or appropriate for the population being tested? If the norms are not available, how will the scores be used to indicate developmental differences? (Often times, evaluations consider relative changes in groups -- intervened vs. control. In such cases the assessments should be extensive enough to demonstrate such group differences. Brief assessment tools with just 5 or 6 items per age category may not be sufficient for capturing treatment related effects.) Will a cut-off score be used to demonstrate “delay”? If so, how will this cut-off point be determined in the population under study?

Another use of assessment of abilities in early child development is to be able to make comparisons across countries, and chart progress over time, as for example in State of the World’s Children by UNICEF. Having globally recognized indicators can facilitate funding and the assessment of progress in a particular area. For example, comparisons in rates of child stunting can now be made globally, since the indicator is agreed on. Sometimes the indicators are extremely difficult to measure (e.g., poverty) and assessment strategies are evolving. There is an effort at the present time by UNICEF and partners to develop a global indicator of a child’s development that could be assessed by parent report during a household survey.

Thorough research relevant to these parameters will narrow the range of tests most suitable for use. After clarifying the purpose for testing, the next issue to determine is the types of assessments to use.

Types of assessments

There are three methods for gathering information on the developmental status of infants and young children: 1) directly testing the child; 2) obtaining ratings or reports of the child’s behaviors or skills by informants, such as parents, usual caregivers or teachers; and 3) observation of the child in daily or structured activities (Grigorenko & Sternberg, 1999; Snow & Van Hemel, 2008). Many tests combine two or more modes of assessment. Each of these methods of individual assessment can be aggregated across groups to create a population-based measure. In this case, it is possible to sample children from a larger pool, rather than test all of

them, which may be particularly useful for large-scale impact evaluations of national programs or interventions.

How well any particular test measures development is important to consider when selecting tests. *Psychometrics* is the area of psychology concerned with evaluating the design and effectiveness of measures to assess psychological characteristics (or domains, such as language, cognitive development, etc.). Psychometric analyses are primarily used to determine the *reliability* and *validity* of an assessment. **Reliability** refers to how consistently a measure produces similar results for a child or group of children with repeated measurements over time. This is based on the assumption that individuals (or groups of individuals) show some stability in how they exhibit the behaviors under evaluation. However, there is typically some variation in scores on successive tests. The reliability of tests can be increased by ensuring that tests are administered uniformly and under conditions where individuals have the capacity to produce their “best” performance. **Validity** refers to the degree to which a measure accurately samples or assesses behaviors or abilities that reflect the underlying concept being tested. For example, do the items included in a language test accurately “tap” a child’s capacity to produce a certain number of words or understand what is being said to them at a given age? (Cueto, Leon, Guerrero, & Munoz, January 2009). The majority of published tests developed in developed countries (USA, UK, EU countries, etc.) have undergone rigorous examination to ensure the assessments are both reliable and valid in the populations in which they were developed; however, reliability and validity need to be determined within each cultural context.

The capacity to measure children accurately (reliably and validly) is enhanced when assessment strategies:

- ❖ **Measure multiple domains** (e.g. language, cognition and socio-emotional development). This provides a more comprehensive assessment of child functioning and can also indicate which domains are or are not affected by an intervention.
- ❖ **Use multiple tests and methodologies to measure both within and across domains** (Grigorenko & Sternberg, 1999). Using two or three measures (e.g. parent report, direct child tests and/or observation) to assess any domain provides a richer developmental profile than any single test could provide, and can then be analyzed in combination. If children can be assessed using two or three methods within the same domain, then the combined results

are more likely to indicate a more accurate, thorough and “true” assessment of that domain. For example, the International Association for Evaluation of Educational Achievement (IEA) Preprimary Project conducted in 15 countries (Montie, Xiang, & Schweinhart, 2006) used both observational measures as well as child administered cognitive and language assessments at age four to examine the impact of pre-school activities on development at age seven. Both types of data were useful in understanding later cognitive and language outcomes. The Turkey Early Education Program (TEEP; (Kagitcibasi, Sunar, & Bekman, 2001)) also used multiple assessments (direct child tests, such as the Stanford-Binet, and parent reports of behavior) to evaluate the effects of a stimulation program on child development outcomes.

Direct Tests

Direct tests assess infants by presenting stimuli such as objects or sounds, to evoke responses or by asking young children to complete tasks or activities, such as stacking blocks, searching for a hidden item, naming objects or climbing stairs. Assessors are usually required to complete training on how to administer and score the test and are often professionals who regularly interact with children in some capacity (e.g., pediatricians, psychologists or teachers). However, other personnel with relevant backgrounds (community health workers, social workers, etc.) can also be trained to conduct these tests. A professional level of training is not necessary for the administration of the tests in an evaluation setting, but a licensed professional would be required to interpret or make a diagnosis for clinical purposes. Examples of direct assessments are the Bayley Scales of Infant Development and the Wechsler Inventories. The pros and cons of this approach are outlined below:

Pros of direct assessment

- ❖ ***Data are gathered first-hand.*** Information gathered via direct assessment (i.e., requiring responses to fairly structured requests by an adult who may or may not be known to the child) is considered to be a “gold standard” measure, because there is no concern about recall bias.
- ❖ ***Data can be very high quality.*** With a highly trained interviewer, data gathered directly from children can be very high quality, and can be less biased than using a parental report. Standardized norms are sometimes available within country and can allow for comparison

with other children. For research purposes, however, changes over time or comparison with a control group do not require the use of standardized norms.

- ❖ ***Many of the “cons” listed below can be overcome with good planning.*** Many of the potential difficulties of direct assessment – and those outlined below -- can be minimized by adapting tests and administrative procedures (see the **Modification and Adaptation** chapter). In addition, scheduling the test during a time when the child is alert; being very familiar with the test so that it moves along seamlessly (that is, without having to fumble around for the proper materials, etc.), and altering the pace of the test in response to the child’s behavior can help elicit optimum test performance (Bradley-Johnson & Johnson, 2007).

Cons of direct assessment

- ❖ ***Young children can be difficult to test.*** The circumstances of direct testing are likely to be unfamiliar to young children – particularly those living in impoverished conditions – and may affect their engagement with the test items. Moreover, young children, as compared to school aged children, may have different motivational styles that affect response to positive feedback and encouragement, or intrinsic desires to do well and please the assessor. Performances on standardized tests may not be indicative of some children’s true abilities (Bracken, 2007). Optimal assessment may be challenged by children’s internal states (hunger, sleepiness) or other behaviors, such as high activity level, distractibility, shyness with adults, low thresholds for frustration and fatigue, fussiness and defiance.
- ❖ ***Testers need a lot of training.*** Accurate assessment of infants is largely dependent upon testers being able to control the infant’s state of arousal, which may be challenged by new stimuli, environments or unfamiliar persons. As a result, assessments may be more indicative of abilities capable of being demonstrated under novel (and perhaps exciting, or upsetting) situations rather than of true mastery in any domain (Snow & Van Hemel, 2008).
- ❖ ***Accuracy depends on the testing demands.*** Tests that include tasks or activities that are new to the child, use unfamiliar words or language structure, require verbal (rather than demonstrative) responses, or require children to choose between qualitative (“best” or “worst”) or quantitative (“more like this” or “less like this”) responses will likely reduce the accuracy of the assessment (Snow & Van Hemel, 2008).

Ratings and Reports

Ratings and reports are scales or checklists completed by informants who know the child well, such as parents/caregivers or teachers. Informants (e.g. mother, father, other caregiver) answer questions about the child's abilities based on what they know of the child, but do not directly assess the child. Ratings and reports can offer information about how children behave in other (i.e., not standardized testing) settings (Snow & Van Hemel, 2008; Squires, Potter, Bricker, & Lamorey, 1998). The rater may simply report about whether a behavior has occurred and how frequently, as in the parent-reported test Ages and Stages Questionnaires (ASQ) (Bricker & Squires, 1999). The rater may also be asked to compare the child with other similar children of the same age. The pros and cons of this approach are outlined below:

Pros of ratings and reports

- ❖ ***Instruments are easy to administer.*** Practically speaking, ratings are usually easy to understand by respondents, requiring minimal instruction or training; are efficient in terms of time and money; tend to be quick and easy to complete, and do not require much time or expertise for scoring and interpretation (Johnson & Marlow, 2006). Parent reports may also be used to estimate stages of development where direct tests cannot be used.
- ❖ ***Parents can become involved with assessment.*** During an assessment, parents also have the opportunity to express concerns that may not otherwise be communicated to pediatricians or other child development professionals.
- ❖ ***Parent ratings correlate well with direct assessments.*** As used widely within the US (Bricker & Squires, 1999; Doig, Macias, Saylor, Craver, & Ingram, 1999; Scarborough, Hebbeler, Simeonsson, & Spiker, 2007) and in some developing countries (Handal, Lozoff, Breilh, & Harlow, 2007; Heo, Squires, & Yovanoff, 2008), there is good evidence that parents across socio-economic levels can provide accurate assessments of children's development as validated by direct child assessments.
- ❖ ***Parent ratings can be adapted for better reliability and validity.*** Behaviors of interest (for which parents are reporting about their children) should be: 1) current and age-appropriate and 2) likely to occur frequently. Furthermore, responses should rely upon recognition rather than recall, and parents should possess the skills and abilities needed to accurately respond to

items. Ensuring that items and response choices are spoken or written in language that is suitable for populations with low literacy rates is also essential.

- ❖ ***With older (3-5 years of age) children, teacher reports may be a valuable source of information.*** Early childhood teachers may be valuable informants of child development as they have multiple, repeated occasions to observe what children can do, how they behave in a variety of situations, and how this compares with other peers of the same age (Snow & Van Hemel, 2008). While research on the use of teacher reports for the purpose of evaluating programs is scarce, there is evidence that early child educators in the US can be trained to reliably use an observation-based rating measure (Bagnato, Smith-Jones, McComb, & Cook-Kilroy, 2002). The Early Development Inventory (discussed below in the **Population vs Individual Testing** section) is a simple teacher rating measure that requires minimal training and appears to be reliable (Janus & Offord, 2007). Moreover, an analysis of its use with some 40,000 children suggests that teachers in certain settings can make unbiased ratings across groups of different children (Guhn, Gadermann, & Zumbo, 2007).

Cons of ratings and reports

- ❖ ***Parents and teachers may inflate scores.*** This trend may be due to social desirability; teachers may also inflate scores if they are used for accountability (Snow & Van Hemel, 2008). It must be emphasized to parent and teacher respondents that it is expected that children will have strengths and weaknesses within and across the domains assessed, and that *only* positive ratings are unlikely. Discussions about normative development and the construction of measures may help parents and teachers to feel more comfortable answering truthfully. If one is comparing ratings across cultures, there may be different tendencies to inflate or deflate scores.
- ❖ ***Parents may not accurately report abilities.*** It is possible that a mother with less education may not be willing or able to report accurately on her child's abilities. If an item is unclear to the respondent, there may be a tendency to simply agree.
- ❖ ***Parents and teachers may have systematically different interpretation of items in different cultures.*** Much care should be taken to ensure items have the same meaning and value cross-culturally. For example, cultural norms about how children should behave at home or

in the classroom (e.g., obedience; not speaking to adults beyond greetings) may affect how children are rated and the intended meanings of the items may be lost.

Observational Measures

Observational measures rely upon a trained observer to document the behaviors of a child. Observational ratings may be completed at home or in an institutional setting (e.g., school or daycare facility), but in all cases, observers must be trained. Three kinds of observational measures are generally used:

- ❖ ***Naturalistic observation.*** Naturalistic observations require the observer to follow the child and observe and record behavior in the normal course of the day. These observations are useful to identify characteristic environments, detect the meaning of behaviors and skills and capacities, and find out the cognitive requirements in a child's life. They are often a valuable component to complement a standardized assessment.
- ❖ ***Sampled observations.*** With sampled observations, specific behaviors can be defined (e.g., caregiver questions a child) and the frequency of these behaviors is observed over a period of time. If the behavior is short and of relatively frequent occurrence (e.g., waving "bye bye") a time-sampling method can be used. If the behavior can vary in length (e.g., a child's crying) then one can assess an event and its duration. For an example, see the International Association for Evaluation of Educational Achievement (IEA) observation system available for download at <http://www.highscope.org/file/Research/international/IEAInstruments/ChildActivitiesObservationSystem.pdf>.
- ❖ ***Structured situations.*** Structured situations are created, and then children are observed in that situation, with a common coding method, to see how they behave. For example, the Strange Situation has been used in many parts of the world to measure a child's attachment to her mother (Ainsworth, 1993). The protocol has the mother leaving and reuniting with her child, and the child's response to the returning mother is coded. Other well known measures are the HOME scale, in which the interviewer observes the caregiver's behavior (Bradley & Corwyn, 2005); a book-reading task in which the mother is asked to read a book with her child (Aboud, 2007); observation of play with specific toys in a controlled situation (Wachs, 1993; Wachs & Desai, 1993; Wachs, Sigman, Bishry, & Moussa, 1992; Wachs, et al., 1992),

and measures of inhibition to respond and infant emotions through measuring reaction to novelty (Leerkes & Crockenberg, 2003; Rubin et al., 2006). All of these have been used in several cultures.

Pros of observational measures

- ❖ ***High on validity.*** Because these measures are based on actual behavior, they are likely to be valid or “true” indicators of typical behavior.
- ❖ ***Measures of behavior in context.*** These measures allow the observer to determine how the child will behave in an identified context (i.e., home or preschool). They may also help the investigator to develop other and more appropriate measures.
- ❖ ***Provide additional or confirmatory information about assessment in any domain.*** These measures may be useful to complement tests and assessments.

Cons of observational measures

- ❖ ***Requires more effort and training.*** To collect this information well requires careful development and examination of the behavioral codes to be used, and extensive training of observers to achieve reliability of coding, and they are more time intensive per child than a test.
- ❖ ***Cultural appropriateness must be determined.*** Some situational assessments may not be appropriate to all contexts, or may be interpreted very differently depending on the culture.
- ❖ ***Coding and analysis.*** Data need to be coded and entered, which may be time consuming if observational codes and definitions are not clearly defined before data are collected.

Screening tests versus assessment of abilities

Screening tests

Screening tests (e.g., the Denver Developmental Screening Test (DDST), Ages and Stages Questionnaires) are brief assessments used to identify – with some degree of certainty – children who are at risk of having developmental problems in one or more domains (Glascoc, 2005). Screens usually include motor, cognitive and language domains, but often do not

measure social-emotional development. They are often used in lieu of ability tests because they are inexpensive, quick and relatively easy to administer, and require minimal time for training. Because screening tests only contain a sample of items per domain (i.e., they do not assess the full range of ability) they do not yield continuous scores, but are used to classify children into categories, such as “Delayed,” “At Risk for Delay” or “Within Normal Limits” for age. These categories have been established for specific populations (typically where the tests were developed), and do not apply to other populations. Screening tests may rely upon direct child testing, parent report or both.

Screening tests (e.g. the Denver) are not diagnostic. These tests can be used, however, in samples where cutoffs have previously been determined to recommend further testing, refer for intervention or to monitor development. Screening tests are not appropriate in samples and situations where cutoffs have not been determined.

Screening tests are often adapted for use in developing countries to detect developmental differences among groups of children because they are easy to implement. However, in most cases, the cutoffs used to classify children into various categories (“Delayed,” “At Risk for Delay” or “Within Normal Limits” for age) in the population where the test was developed have **not been verified for the population under study**. Therefore, the use of screening tests in countries where no population-based cutoffs have been established to determine such classifications should be **limited to examining how one group of children performs on the screening test relative to another group of children**. In these situations, the screening tests are used as a ‘gauge’ of relevant development rather than as a screen to recommend further testing or services. Cutoffs used in one population to classify children as ‘delayed’ or ‘normal’ should never be applied another population; however, screening tests can be useful for examining developmental differences in groups of children. Cut-offs appropriate for the population under study should be determined if it is desired to identify children as developing normally or delayed relative to other children in the same population, but this should only be done where the classification is useful for recommending enrollment in programs or services.

Ability tests

Ability tests include those designed to assess the maximum skill level for a child at any given age. These tests are often direct child assessments (e.g., the Bayley Scales of Infant Development), but can also be parent or other informant report by way of milestones or language checklists (Lansdown, et al., 1995; Stoltzfus, et al., 2001). Ability assessments provide detailed, comprehensive information on children's developmental levels within domains and as a summary across domains. Scores are frequently standardized and can be used to compute developmental quotients (developmental age/chronological age x 100), or DQs. The main advantage of ability tests that produce continuous scores is that scores can be used to compare children's developmental levels with more precision, and scores may be more sensitive to treatment effects, as compared to screening tests, because they measure differences at the upper end of the distribution as well as the lower. Younger children's scores (under age three) are typically labeled as Developmental Quotients, as they may still change, whereas for older children the scores are called Intelligence Quotients (IQs) as they become more predictive of future development. Some tests are diagnostic, assessing specific skills such as communication, and can be used to recommend and design types of remedial assistance. While ability tests can be time-consuming and require a high degree of training to conduct, they provide flexibility in how scores can be used (that is, as raw scores, DQs, IQs, or with cutoffs for determining delay as specified within a population).

Population versus individual-based testing

The purpose of the majority of child development assessments is to measure how individual children progress as a result of a health, care-giving or educational intervention, or to relate performance on one test with another. Population level measures, on the other hand, are used to compare a group of children (such as within a classroom or a school) to other groups of children. They are also used to track overall progress, as noted above; to inform policy decisions, and to inform about appropriate planning of interventions. This method encourages a focus on the context of children's abilities, and community level factors, and reduces the risk of using tests to categorize children and in some cases, even stigmatize them. However, it is important to note that any test could be used as a population measure by aggregating across

groups. A review of the benefits of using population level measures is available (Mustard & Young, 2007).

Rationale for population-based tests

The rationale for constructing a population rather than an individual measure depends more on how it is used rather than the specific assessment method or the validation strategies. Any measurement can be used to provide group-level or population-level data, similar to that in a census (Janus & Offord, 2007). “Population-level community reporting theoretically can be achieved by aggregating any measurement available for all individuals in the community, or a representative sample, similar to the way census reporting is carried out” (p. 10) (Janus & Offord, 2007). The preference for a population-level assessment is based on a population-based model of health that argues that small problems over a large number of individuals will contribute more to the burden of ill health than severe problems in a minority of people (Janus & Offord, 2007). The use of an assessment as a measure of “population health” also assumes the importance of community strengths and weaknesses, and assesses the value of community-oriented interventions (e.g., providing local libraries) that is not captured by individual assessments. For example, each school could be considered a category, and schools could be compared on the percent of children at risk. The intervention, such as financial resources, could be distributed to schools on the basis of the school-level variables such as the percent of children at risk. It is also possible that not all children are assessed, but that a group could be sampled from the larger population.

EDI: example of a population-based assessment

An example of a measure designed to be population-based is the Early Development Inventory (EDI), a teacher rating of children’s readiness for first grade, assessed during kindergarten. It differs from many other tests developed to measure children’s maturational or experiential readiness for school (Janus & Offord, 2007). In their review of 7 of these instruments, Janus and Offord (2007) concluded that although some are reasonably predictive of school success, they have to be administered by a professional, they are not cost-effective, nor do they measure all relevant domains (e.g., social emotional development was missing). To fill this gap, Janus and Offord developed a teacher-rating scale that can help assess children’s school readiness at a much lower cost.

The Early Development Inventory is a set of questions (initially 103, but there are shorter versions) that a teacher can use for rating an individual child (Brinkman, et al., 2007; Janus & Offord, 2007). The questions cover five domains: Physical Health and Well-Being, Emotional Maturity, Social Competence, Language and Cognitive Development, Communication Skills and General Knowledge. From the instrument, one can rate whether each child is vulnerable on each of the five dimensions. The ratings were found to be associated with other measures of cognitive and socio-emotional development (teacher ratings on other measures, direct tests, and parent ratings) and thus had reasonable construct validity in both Canada and Australia (Brinkman, et al., 2007; Janus & Offord, 2007). These associations were compared using both a continuous scale and a dichotomous measure of vulnerability. Associations with other teacher ratings and with tests were reasonably high, and higher than with parent ratings.

Survey-based measures

Because there is no global indicator for early child development, making governments and policy makers aware of the importance of development during the first five years of life has been difficult. Recently, there have been several attempts to develop survey-based assessments of children's development using a rating system like the EDI, but based on parent ratings rather than teacher ratings. After a careful examination of the literature, and pilot testing in two countries, UNICEF developed an 18 item simple version of the EDI that asks parents to rate their child's behavior in five domains of development. This new measure will be included in the next round of the Multiple Indicator Cluster Survey (MICS4).

Pros of population-level measurements (or using an individual test at a population level)

- ❖ ***Community-level focus and planning is possible.*** The advantages of a community or population-level measurement are seen in the planning process, as it allows the planner “to identify inequities related to characteristics which may be remediable by appropriate policies. The EDI is the first tool for which community aggregation is feasible due to low cost, relevance in covering the 5 developmental domains, and proven psychometric properties” (Janus & Offord, 2007).
- ❖ ***Protects the individual child from being categorized or stigmatized as “low” based on a test score.***

- ❖ *Far less expensive data collection method than individual testing.*
- ❖ *The measurement technique may be useful for survey-based research and asking parents for their reports of child behavior.*

Cons of population-level measurements

- ❖ *Teachers may show systematic biases in their ratings.* This is a particular issue when measures are compared across cultures, or across children varying in age level. When teachers are less sensitive to bias, cultural categories such as income, race, or gender may influence ratings.
- ❖ *Ratings will vary by age.* When children of many different ages are in the same group or classroom – as is common in some developing countries -- ratings will most likely vary by age, but the tests are not age-adjusted.
- ❖ *Sampling is difficult.* When not all children are in the group measured (e.g., children attending preschool) it is more difficult to obtain a truly random sample. If the sampling is done in a school or preschool setting, non-attenders are not included.

Ethical risks and responsibilities in assessing young children

Beyond deciding which instruments to use, assessment teams must also be cognizant of the risks and responsibilities associated with the assessment of young children.

- ❖ *All assessment protocols must be reviewed and approved by an ethical review board.* Many universities and non-governmental organizations have Institutional Review Boards (IRBs) that fulfill this role. If investigators in the United States or another developed country are working with researchers in a low- or middle-income country, it is generally not sufficient to have approval just from the developed-country-investigator's home institution. Wherever possible, it is essential to have protocols and permission forms reviewed by an IRB in the country where the study is taking place. In the case where the person administering a child's assessment is not affiliated with an organization that has an ethical review board, an external institutional review can be sought. For example, the Western Institutional Review Board is an organization fully accredited by the Association for the Accreditation of Human Research

Protection Programs, Inc., which will review and approve study protocols involving human subjects. The Office for Human Research Protection of the United States Department of Health and Human Services (<http://www.hhs.gov/ohrp/>) mandates that all research funded by the US-based National Institutes of Health must be approved by an IRB before it can receive any federal funding.

- ❖ *Accuracy and validity are extremely important, especially where assessments are used to identify children with delays (within a population where such cutoffs have been determined).* Non-professionals administering tests must be well trained and understand the objectives of testing when using screening and ability tests as the failure to identify children who are delayed by local standards (false negatives) may result in children not receiving needed services or interventions. On the other hand, wrongly classifying children as delayed (false positives) within the population can cause needless distress and worry for parents (Fyro & Bodegard, 1987; Tluczek, et.al., 1992). Moreover, being labeled as delayed according to local norms of development – even if later repudiated – can follow a child, possibly affecting self-perceptions as well as how a child is perceived and treated by peers, teachers and the community. Using a screening test out of context with inappropriate cut-offs for a given population is not ethically justified.
- ❖ Follow-up should be mandatory. Testing should be carried out with the intent that appropriate follow-up, such as referrals for services or further monitoring, will be provided (Glascoe, 2005; Ryzd, et al., 2005; Snow & Van Hemel, 2008). While this can be difficult to accomplish in developing countries, consideration of how the testing will benefit individuals or communities is imperative. It is the clinical team’s responsibility to ensure that measurement yields the most accurate information possible about a child’s development; the local team and community understand the meaning and utility of the test results, and that the results are only used in ways that benefit the child.

Constraints to consider

There may be practical and logistical issues that will affect the selection of tests. Intervention teams may find it necessary to consider which tests suit the project best and are feasible given constraints of factors such as:

- ❖ **Budget.** Many standardized tests are prohibitively expensive for use in large-scale studies. For example, the Bayley Scales cost ~\$1,000.00 per test kit per interviewer conducting assessments and about \$1.50 per child assessed using the materials, while the Ages and Stages Questionnaires cost about \$200.00 per interviewer with no additional cost per child. Many testing companies, however, will offer a research discount of 20-50%; these discounts often take 6-8 weeks to obtain. Tests vary in terms of how much interviewer time they need. For instance, the Bayley Scales can take 30-60 minutes to administer, and interviewers must be paid for training time in addition to administration time. Furthermore, the Bayley test usually needs extensive amounts of pilot testing and adaptation of materials, which can add to the expense.
- ❖ **Copyright issues:** Most of the tests developed and licensed in the developing world (e.g. Bayley Scales, Denver Development Tests, Woodcock-Johnson, etc) are strictly protected by copyrights. In many cases, a licensed psychologist is the only person that can purchase the tests from the publishing companies. Copyright laws prohibit any use of the tests (including photocopying) without explicit permission or purchase. Furthermore, translation is not allowed without approval from the legal department of the publishing companies.
- ❖ **Time allocated for testing:** Direct child tests will likely take from 20-60 minutes; screening or parent report tests may take 30 minutes or less. Direct testing for infants and toddlers should take account of the fact that children may tire and become hungry during the course of testing and time should be allowed for this possibility.
- ❖ **Training** (capacity for administration): Some tests require considerable time – one or two months – to adequately train and standardize testers. Standard good practice is to obtain reliability of a new tester with a “gold standard” tester.
- ❖ **Test setting:** field vs. clinical or lab testing. Children may be very uncomfortable being tested in locations that are unfamiliar and perform poorly, so home or school testing can be preferable. The drawback of home or school testing is that the testing environment will vary according to characteristics that could, in themselves, affect performance (e.g. lighting, noise, seating). For this reason, it is critical to make the environments as homogenous as possible to minimize distractions and maximize consistency. For example, the research team could carry a folding table and two chairs, and only test during daylight hours so that the testing

environment itself is identical even if the location is not. Similarly, the research team could include someone whose job it was to maintain a quiet atmosphere, but making sure there were no observers or other distracting onlookers. If an unfamiliar testing environment is to be used, every effort should be made to make the place cozy and comfortable.

- ❖ ***Capacity of respondent:*** In the case of using rating assessments, the ability of respondent (e.g. parents or teachers, doctors, etc.) to report accurately on children or rate children, is critical to the success of the rating assessments. Before using rating measures, this issue should be considered.
- ❖ ***Language and cultural differences:*** When tests are adapted to different cultures, there may be considerable differences in the difficulty and meaning of items. Care must be taken that tests are thoroughly evaluated in each language (within and between samples), or at a minimum, that the test is back-translated and reviewed by skilled bilingual interpreters. For language development, it is crucial to have a detailed understanding of the typical progression in native speakers. In the absence of a detailed linguistic ethnography, a literal (word-for-word) translation may ask about irrelevant words or words that do not capture the intended meaning as used in the original test. The testing situation should support children's best performance within the cultural context (e.g., reducing the requirements for verbal expression in a culture that does not support it). Test materials should be familiar within the culture or adapted so that they are understandable. The language of assessment should be the child's native language. If cultural knowledge of the test items is limited, a reduced set of items might be used. Finally, timed items might be problematic for a child who is unfamiliar with the importance of a speedy response, and may not provide meaningful or valid results.
- ❖ ***Materials:*** Many of the commonly used tests have pictures or figurines, objects like bells or staircases, or materials like brightly colored plastics, which are unfamiliar to many children living in developing countries, especially those in rural areas. These often need to be replaced with locally produced materials. Similarly the text or pictures may describe practices (like sitting around a table having a meal together) that are not part of the local culture, and will need to be replaced. Many of the pictures in the Peabody or KABC-II, for instance, depict objects that are not available in rural communities in developing countries.

Such necessary adaptations may be costly in time or money and may constitute constraints in selecting instruments.

CHAPTER 4: MODIFICATION, ADAPTATION & STANDARDIZATION OF EXISTING TESTS

Introduction

Children are embedded in cultural systems from birth. Therefore, almost all developmental capabilities are in some way affected by the opportunities children have to develop their skills, the attitudes and beliefs of their caregivers, and their caregivers' expectations for healthy development. Some cultural practices may have more substantial implications for development than others. Although even the emergence of canalized abilities, or those that all normally developing humans eventually acquire (such as walking and talking), are affected by culturally dictated child-rearing practices, it may ultimately be of little consequence. For example, children who are carried on their mothers' backs tend to walk at different ages than children who spend more time moving independently, but when children learn to walk appears to have little bearing on their future development. Cultural practices around literacy (such as a belief that boys are more capable of learning to read than girls), however, may strongly affect development through avenues that are not readily apparent to evaluators, and in turn may affect the impact of an intervention on children's outcomes even when the intervention is working properly. Therefore, when selecting measures to use in each country, prevailing cultural beliefs and practices should be carefully documented to aid in the interpretation of the data and the conclusions regarding the impact of any intervention on children's development.

Norming and milestones

Many tests, especially tests of cognitive development, have been "normed," meaning that test producers have collected enough data, usually from developed countries, to draw a normal curve of scores; the USA, Canada, the UK and Australia usually have their own norms. These norms allow one to compare children's performance across different ages.

Yet these norms, while indicating how far children in the developing world deviate from children in the norming country, may not be useful in assessing normative development for

children in diverse contexts. Instead, it may be useful to view the first task of assessing child development in the developing world as collecting information on milestones, or when children of each population group tend to display certain developmental achievements. Collecting this information will allow a more culturally appropriate comparison of the developmental status of children who are receiving an intervention and their peers, and will also supply information on the emergence of skills in diverse contexts.

Validity of measures

While many language and cognitive-developmental assessments have been used in developing world contexts, to date, there is little research that has validated these assessments by examining longitudinal relations between scores obtained prior to school entry and children's performance in school. As a result, while we know that scores on these assessments predict children's school performance in the United States, much less is known about the long-term predictability of these assessments in the developing world. It may be useful to expand the scope of assessments to include measures of executive function (inhibition, working memory, attention), because they assess children's capacity to adapt and respond to new situations or conditions in order to achieve a particular goal. The literature suggests executive function may be more predictive of children's academic functioning across diverse contexts than knowledge alone.

Lack of information on normative developmental milestones

There are only a few published studies offering a description of the achievement of children's developmental milestones in developing countries (see **Creating new tests** below). Recently, WHO collected information on 6 motor milestones from well-nourished populations in 5 countries, providing information on normative attainment of various skills (WHO, 2006b). Unfortunately, no other developmental dimensions were assessed. In the 1990's, a team lead by the WHO assessed performance on milestones of over 20,000 children in India, China, and Thailand (Lansdown, et al., 1995). Several other research teams are attempting to do this (e.g. Kilifi Developmental Inventory; Grover-Counter Test; ECCD Philippines; ICMR Milestones – see Appendices and attached Table). For infants, this information is critical for assessing children's development due to the substantial variation in “normal” development among infants in different cultures. Therefore, for very young children, capitalizing on the opportunity to

collect descriptive information on when children are achieving developmental milestones may substantially improve the ability to determine whether programs are having desired effects in a particular cultural context.

Modification and adaptation

No test is “culture-free;” however, many assessment teams choose to use existing tests that have already been shown to be reliable and valid assessments in the context where they were developed rather than develop tests anew. Using existing tests requires careful adaptation and modification. Adaptation refers to processes (including translation, item modification) researchers undertake to reduce systematic *bias* or *error* in test scores that can occur when applying a test in a culture other than the one in which it was developed. There are three main types of bias.

- ❖ ***Construct bias*** occurs when the instrument does not measure a construct (intelligence, social-emotional development) the same way in both cultures. This may be due to differences in the definition of the construct; variability in the measurable behaviors and skills that represent the construct, or inadequate coverage (too few items or domains) to sufficiently assess the construct.
- ❖ ***Method bias*** occurs when the administration or procedures of the test -- the use of unfamiliar stimuli (blocks, puzzles) or unfamiliar response formats (scales, multiple choice) – differentially affect the scores of groups being tested.
- ❖ ***Item bias*** occurs when individual test items do not measure the same way across groups. Sources of item bias include poor translation and culturally inappropriate content (Van De Vijver & Hambleton, 1996).

These biases threaten the validity of tests’ capacity to produce “true” score of children’s abilities (Peña, 2007); bias can be reduced, however, by examining how *equivalent* the adapted test is to the original. There are four types of equivalencies that can be considered (Peña, 2007):

- ❖ ***Linguistic equivalence***, or is the translation accurate?

- ❖ *Functional equivalence*, or do the instructions and items have the same functional meaning (i.e., do they get at the same idea and produce the desired response) in the two cultures?
- ❖ *Cultural equivalence*, or do the instructions and items have the same relevance or meaning?
- ❖ *Metric equivalence*, or do the items have the same level of difficulty?

While the International Testing Committee (ITC) has published broad guidelines concerning the use and adaptation of psychological and educational tests internationally (<http://www.intestcom.org/Guidelines/guidelines+for+test+use.php>; <http://www.intestcom.org/Guidelines/test+adaptation.php>), there are no universally recognized minimum standards for what child test *adaptation* should entail (Carter, et al., 2005; Malda, et al., 2008; Peña, 2007; van Widenfelt, Treffers, de Beurs, Siebelink, & Koudijs, 2005). Several aspects of the adaptation process are repeatedly cited, however, as indispensable to producing a valid adaptation (Carter, et al., 2005; Hambleton & Patsula, 1998; Malda, et al., 2008; van Widenfelt, et al., 2005). These include translation; the selection and adaptation of culturally sensitive content; ensuring test stimuli are culturally relevant; and identifying presentation, administration and scoring procedures that maximally reduce cultural-based differences in response or performance (Bracken & Barona, 1991; Mwamwenda & Mwamwenda, 1989). A discussion of these aspects is included below. Examples based on some of the authors' experiences adapting the Ages and Stages Questionnaires (ASQ) (Bricker and Squires, 1999) in various countries are also provided.

Preparatory work for test adaptation

- ❖ *Involve local professionals.* The team adapting the measures must include local professionals -- psychologists, social and community health workers, early child education teachers, doctors -- who work with young children and their families (Malda, et al., 2008). Include professionals who have experience with rural or urban children according to the population you will test. These team members will be essential to gathering both general and specialized information on linguistic, cultural and technical aspects of the test adaptation.
- ❖ *Test items within the community.* Engaging small groups of local key informants (e.g., parents, teachers and others working with young children) is an ideal way for gathering information on the test content and procedures. This process includes using a somewhat structured interview to ask groups of respondents to re-phrase the items and responses to

ensure they are understood accurately. Respondents should also be asked which response they would select and explain how they arrived at that answer (Alaimo, Olson, & Frongillo, 1999). This interview technique can also be done to get feedback on the test stimuli (e.g., “What does this picture mean to you?”) as well as various response set formats (multiple choice, scales, etc.) to assess their suitability.

Steps for successful test adaptation

❖ ***Produce an accurate translation (linguistic, functional equivalence).*** Ideally, it is recommended that the translation process include 2-4 individuals who are bilingual and bicultural. Multiple team members enable identification of problematic translations (Solarsh & Alant, 2006; van Widenfelt, et al., 2005). While it is generally preferable to keep the translation as close as possible to the original test, it should also be kept in mind that word-for-word translations may not retain the original meaning of an instruction or item (van Widenfelt, et al., 2005). In such cases, the team needs to develop and test alternative translations to identify the one that best captures the meaning of the original phrase. For example, the piloting of a bilingual language test used with 4-6 year olds found that instructions to Spanish speakers to “Describe...” a particular object was equivalent to (i.e., got the most similar responses as) the English instructions, “Tell me three things about...” a particular object (Peña, 2007). Similarly, a translated and adapted version of the Denver Developmental Screening Test used in Costa Rica altered the instruction “draw a man” to “draw a doll” to produce a response most similar to the original item (Howard & De Salazar, 1984). It is also possible that literal translations will result in language that is too complex for the respondents. In populations where literacy levels are low, exact translations may need to be simplified in order to increase respondents’ comprehension of the test (Peña, 2007). Translations should strive to be at the most basic level possible.

The steps involved in producing an accurate translation include (Solarsh & Alant, 2006):

- 1) Translation and back-translation (by two different individuals) of all test instructions and materials;
- 2) Review and comparison of back-translated test with original language test;
- 3) Corrections of the translated version as necessary

- 4) Confirmation of the translation by another bilingual adult living in the community
- 5) Trying out the instructions to the child with children in the target community. Often when there is local variation in a language, young children are only aware of the local words. Also, children may misunderstand instructions that do not present any difficulty for adults.

The team should also check for poor or incomplete translations that may occur when a translator is unfamiliar with the underlying concepts of the items or tests. For example, a Mexican Spanish translation of the (English) Ages and Stages Questionnaires (ASQ) item, “When playing with sounds, does your baby make low-pitched noises?” resulted in “When you play with your baby, does s/he make low-pitched noises?” changing the meaning of the original item and had to be adjusted.

❖ *Adapt test content to the local context (functional, cultural and metric equivalence).* The test content may need to be altered to ensure items elicit behaviors or responses similarly across cultures (Peña, 2007). To accomplish this, the ideas or situations expressed in the item must be relevant, easily recognized and readily understood in the local context, and also match the difficulty level of the original item (Solarsh & Alant, 2006). For example, in adapting the ASQ for use in Mexico, an item about whether a child asks a caregiver to wind up a toy was replaced with whether a child asks a caregiver to open something (such as a bottle) or peel something (piece of fruit). In addition, test stimuli (balls, blocks, dolls, etc.) may need to be replaced with objects that are found locally, and pictures and drawings should depict people, houses, trees, animals, etc., that are familiar to the setting (Carter, et al., 2005). Where child development tests require caregiver responses, consideration should be given to cultural norms that may affect how adults understand and answer questions. Where formal education is not universal, caregivers may lack experience reflecting on their thoughts or making relative comparisons. In cultures where thoughts are not distinguished from what is “real” and observed, caregivers may not be able to respond to items asking them to imagine hypothetical situations or make speculations (Greenfield, 1997).

Response sets may also need to be changed to make certain that the response choices are unambiguous and represent the desired complexity. For example, multiple choice tests should include possible responses that are similar in difficulty to the originals, ensuring that there is one clearly correct answer but that it is not too obviously correct. Gradient scales

using numbers or phrases may need to be substituted with illustrations or objects that represent the response options, or using hand gestures (to indicate more or less).

In some cases, there may not be suitable cultural equivalence of an item for the age being tested. In our experience adapting the ASQ, we found children do not frequently use forks in Peru, Indonesia, and Tanzania, and as there was a previous item asking about use of a spoon, no suitable substitute items could be found. Thus, this item was dropped from the test. This does not mean, however, that shortening tests at will is appropriate. Some assessment teams may be tempted to abbreviate standardized tests during adaptation to better suit the project demands (large samples; limited time and resources). Snow et al (2008) warn against this, as shortening a measure may threaten its reliability, validity and equivalence with the original test (Snow & Van Hemel, 2008).

- ❖ ***Adapt the administration procedures (functional and cultural equivalence).*** Tests standardized in the USA or UK typically identify the range of items to be used with children of a particular age. These age-specific item sets reflect how items worked in the country in which they were developed and may not be appropriate in other countries. Adaptation teams should explore which set of items most accurately assesses development at particular ages by piloting a larger range of items (i.e., from younger and older item sets) in a representative sample. Re-ordering of individual items may also prove necessary, based on their performance in the piloting. In Indonesia, a child's use of the pronouns "I" and "me" occurs at later age than when asked in the ASQ administration. Further testing would have to be done to determine at what ages items about pronoun use should be asked.

Many adults and children will be unfamiliar with "test" taking, and therefore the very situation of being asked questions and responding to a stranger will be foreign, which could interfere with test performance. In addition, women and children may be very shy. Some suggestions for overcoming these issues are:

- 1) ***Tester:*** The tester should understand the test materials well, be of the community and fluent in the language spoken by the respondent. An open, engaging, non-judgmental approach toward the testing will be less likely to intimidate the respondent, especially young children. It may be important to alter the pace according to the child and culture. Special training is needed to make sure that assessors can encourage a child to try to

answer difficult questions. The training of the assessors needs to focus on shyness of children in cultures where children are not encouraged to speak to unfamiliar adults or to voice opinions in the presence of adults, and how to deal with that situation. Similarly, the manuals associated with assessment tests need to deal with standardized procedures for difficult-to-test children.

- 2) ***Test environment:*** There are two issues to consider. In the absence of a standardized setting, testers should attempt to simulate ideal testing conditions (fairly quiet place that provides some privacy to respondents; a place with sufficient light and space to complete all items) as best possible across all test administrations. The second issue concerns creating a friendly, non-threatening atmosphere. This begins with ensuring the child is accompanied by the caregiver or other familiar adult throughout the testing, and may include adapting procedures so that: the tester sits next to and at the same level as the respondent (Baddeley, Meeks Gardner, & Grantham-McGregor, 1995); not asking questions directly to the child if culturally inappropriate (Snow & Van Hemel, 2008); spending additional time chatting with the respondents or household members to establish rapport, and providing toys or materials for child to play with before beginning the test.
- 3) ***Test procedures:*** The instructions or procedures may need to be altered in order to elicit the best performance possible from the respondent. These changes should be discussed with the local team and may include: allowing extra time for a child to become sensitized to test stimuli prior to administering an item using the stimuli; allowing additional practice trials for items that contain unfamiliar stimuli or activities, such as engaging in grouping or sorting tasks or working puzzles; allowing extra time than recommended in the original test for completion of timed tasks (understanding about the importance of time should be explored as this may differ cross-culturally); the types and frequency of praise, encouragement, feedback or probes used throughout the testing. It is important to explore which types of praise and encouragement (words or gestures or both) work best with the target child. There should be praise at the beginning of each test or section, tapering off to active interested attention. If there is verbal praise after each response, then children notice when the tester does not praise. The effectiveness of probes such as "Tell me more" should also be explored with both children and adults to ensure their use

has the desired effect of increasing test performance (Peña, 2007). Additional, clarifying instructions may also be required.

- ❖ ***Conduct a pilot test.*** Tests should also be administered to a small pilot sample representative of the population where the test will be used. A debriefing with respondents (adults) after the pilot testing can also provide additional information on aspects of the test procedures. The psychologists involved with the adaptation should examine the psychometric properties of the test. These analyses include determining the internal consistency of the measure (i.e., how well the items work individually and together as a test), with a test such as Cronbach's alpha (Cronbach, 1951); examining whether expected age-related differences are evident; and ensuring the items show good variability (e.g., not all children got an item correct or wrong). The test may then require several adaptations and re-tests to reach the best tool.
- ❖ ***Allow time and resources for iterative adaptation and testing of the tool.*** The adaptation of tests is likely to require multiple "rounds" for each step outlined above to ensure the test is valid. Ideally, researchers should allow at least three months for completing this process from start to finish. The time needed will vary by number of tests being adapted, the access to samples similar to those who will be examined, the availability of adaptation team members, etc.

CHAPTER 5: CREATION OF NEW TESTS

Introduction

Rather than adapting an existing test, research teams occasionally elect to create their own tests. This may be done when previously adapted measures are not available, or when copyrighted tests are too expensive to use. The great advantage of creating local tests is that they can be tailored to the local context. Often, this process involves compiling items from existing tests that include items known or believed to validly measure concepts in the population under study (see Gladstone, et al., 2008; Holding, et al., 2004; Stoltzfus, et al., 2001).

Alternatively, some researchers may be interested in identifying and measuring locally defined concepts of child competence (Lansdown, et al., 1995). Before undertaking the development of such tests, there should be a clear idea of how this measure would provide information that would discriminate between groups of children under study (i.e., intervened vs. control) and how these measures would relate to other intervention goals (that is, school achievement or adult productivity).

The development of any new test requires employing the procedures outlined above for modifying and adapting tests, as well as a more detailed examination of how the new test works (more detail is provided below). Ultimately, it should be demonstrated that scores on the new instrument measures the domains similarly to other assessments (if possible) (Hambleton & Patsula, 1998) or correlate with factors (e.g., physical growth, caregiving practices, maternal education, SES) known to be predictive of outcomes being measured.

Requirements for creating a new test

- ❖ *Involvement of an inter-disciplinary research team*, including bilingual psychologists who are able to ensure a psychometrically sound process is employed in the development of the test, and (if different) local psychologists who are able to provide insight to the constructs being defined and instrumentalized.
- ❖ *An adequate representative sample for testing items and test cohesion*. New assessments must be piloted with a sample similar in age, sex, ethnicity and socio-economic status as the target population.

- ❖ *Detailed analyses of the instrument’s psychometric properties*, so that thorough examination of how the measure “works” can be made. This includes:
 - ❖ Does the instrument adequately cover the entire domain or concept intended to be measured? If a test is measuring language, for example, do items address both receptive and expressive language abilities?
 - ❖ Are the items ordered to reflect age-related progression in the domain under study?
 - ❖ Is the test reliable, or do the items assess the concept the same way over time (test-retest scores are highly correlated)?
 - ❖ Do the items measure the same way in different groups (e.g., poor vs. less poor) of children? (That is, there should not be items on the test that only children of higher SES or living in a rural region, for example, can pass.)
 - ❖ Do scores on the scale vary meaningfully by subgroups of children in the sample? If it is of interest to create a national tool, is the pilot sample nationally representative and of sufficient number to detect developmental differences?
- ❖ *Development of norms or standards that represent typical development* in the population under study so that recommendations for services or meaningful interventions can be made. This can be much more demanding in time, effort and resources and required expertise, and the resulting measure may not be comparable with other measures of similar constructs.

There are many examples of new tests developed for a particular cultural framework (see Appendices). In each case, the tests were developed in order to be appropriate for the cultural context or specific assessment need.

One elegant example of this process is the study undertaken by WHO in the 1990s (referred to above) to produce culturally relevant developmental checklists (for screening) for use in the home, community or in primary care centers (Lansdown, et al., 1995). The tests were developed in several phases in China, India and Thailand. A total of 28,115 children 0-6 years of age were tested during the process of creating and selecting the motor and mental milestones. While each country maintained longer versions, 13-19 key milestones were ultimately selected by each country for use in health clinic or community centers. The inclusion of overlapping

behaviors enabled the authors to create norms (median age at attainment) for comparison within and across sample sites. Examples include “sits” (range 5.4 months in Thailand to 6.9 months in rural China); “uses cup” (9.5 months in Thailand to 35.4 months in urban India); “says one word” (9.7 months in urban India to 15.0 months in rural India). Each country also included culture specific items, such as “use of chopsticks with small foods” (31-33 months China) and “ties sticks together with string” (45.7 months) and “carries wooden block on head for 5 steps” (45-47 months India).

Some other examples (by region) of new country-specific tests that have been developed include:

❖ Africa:

- The Kilifi Developmental Inventory (Abubakar, Holding, van Baar, Newton, & van de Vijver, 2008; Abubakar, et al., 2007; Abubakar, Van de Vijver, et al., 2008) was developed to assess psychomotor development in a resource-limited setting. The Kilifi is a continuous measure and was originally designed to assess effects of malaria on functioning.
- The Grover –Counter Scale of Cognitive Development (Sebate, 2000) (see <http://www.hsrc.ac.za/ECD-Measure-158.phtml>) was developed in South Africa to assess the level of cognitive functioning of children 3-10 years of age with impaired verbal skills, whether receptive, expressive, or both. It is language-free and based on Piagetian concepts of development. This test was designed to facilitate diagnosis and treatment for mentally handicapped, but may also be used in populations where many languages are represented or where children are very shy.
- In Malawi, a developmental test was developed by combining items from the Denver Developmental Scales (Frankenburg, 1985; Frankenburg, Dodds, Archer, Shapiro, & Bresnick, 1992), the Griffiths’ test (Griffiths, 1984) and some new items drawn from culturally sanctioned behaviors (Gladstone, et al., 2008).
- The Parent Rating Scales of Motor and Language Development (Stoltzfus, et al., 2001) measures gross motor and language milestones via parent report for children 6-59 months of age; used in Tanzania (and Nepal).
- For further reference, also see <http://www.hsrc.ac.za/ECD-Measures.phtml>.

- ❖ Asia (also see above example):
 - In India, the ICMR Psychosocial Developmental Screening Test has been used both as a screening instrument and as a tool for assessing group differences in intervention research (Vazir & Kashinath, 1999).
 - The Cambodian Developmental Assessment Test (UNICEF, Cambodia) measures the level of cognitive, social, motor, and academic development for program evaluation based on country-specific standards.
- ❖ Latin America:
 - Test de Desarrollo Psicomotora (TEPSI; (Hausler & Marchant, 1980)), developed in Chile, evaluates child development in three basic areas--motor function, coordination, and language--by observing behavior in certain situations set up by the examiner.
 - Escala de Evaluacion del Desorrollo Psicomotor (EEDP) developed in Chile (Rodriquez, 1996) is a screening measure of language, social, coordination, and gross motor skills. Norms and cutoffs have been determined to classify children as normal, at-risk, and delayed.
 - Escala Argentina de Inteligencia Sensorimotriz (EAIS) (Oiberman, 2005, 2006) is a diagnostic qualitative measure of practical intelligence in the sensory-motor period. The test is based on the observation of the child's behavior in a variety of tasks.
- ❖ Multi-national:
 - The International Association for Evaluation of Educational Achievement (IEA) developed cross-national tests of language and cognitive development, as well as child observation tools, for use in 15 different countries with children at age 4 and 7 years of age (Montie, et al., 2006).

The “Standards” approach

Another approach to child assessment is for a country to develop a set of standards or expectations about what every child should know and be able to do at a certain age (often four

years, before the child enters school) (Kagan & Britto, 2005). These standards, or desired results, can be linked with program standards for a health or child care center program, resulting in a system of childhood assessment in which the expectations for children and the expectations for programs are aligned for maximum effectiveness. For example if there is a standard that children should be able to understand the concept of sequence by age four, then the program should be assessed in terms of its ability to provide opportunities for learning how to sequence.

In developing standards for early learning and development (ELDS), domains are defined, and within each domain, a set of standards or goals for children are established. For each standard, a set of specific objectives are outlined for the age level, and indicators for each are specified. Indicators are often broad descriptions of behaviors and lack the specificity needed to develop a test, but are intended to help a teacher observe a behavior.

The process of developing national-level standards can be of value for a country, as it brings all stakeholders together and makes them define goals and actions for children – but the process takes time. The advantage of countries' development of their own standards is that they cover items and domains important to the country. If governments have not developed their own child and program standards they may find it more convenient to simply adopt standards from another country. This could lead to inappropriate standards unless modified for the setting.

Therefore a major effort, beginning in 2003 and led in part by UNICEF, has been to help countries define what they expect children of various age groups to know and be able to do (Kagan, Britto, & Engle, 2005). Country teams (experts, policy makers, teachers and families) first define the most appropriate domains for their country, possible sub-domains, and the age groups for which they wish to define standards. The next step is for the country to develop a set of standards, or expectations for learning, that are appropriate to their cultural context. Standards are statements that specify an expectation for achievement of skills or knowledge. Within each standard are several indicators which can be used to assess the standard. Domains may have sub-domains defined as well, a standard and several indicators. A complete set of standards would include suggestions for activities for achieving these standards.

In sum, for each domain or sub-domain of development (e.g., cognitive, language, social, or physical) there are a set of statements of what children should be able to do, and a series of indicators that a defined percentage of children should be able to do by a certain age.

Researchers tend to use the 50% passing rate, although others use a 75% passing rate in order to be sure that children are not mislabeled as “slow” when they are not yet there.

Two examples from Vietnam for children 5-6 are shown below. One can note that these performance indicators are not yet specific enough for testing. These indicators are often used to help teachers of young children to plan curriculum, improve teaching, and develop awareness of children’s skills. In situations where they will be used for assessment in a systematic way, these performance indicators must be much more carefully defined and specified.

<p>DOMAIN 3. COGNITIVE DEVELOPMENT and APPROACHES TO LEARNING</p> <p>Sub-domain 3.1. Cognitive Development</p>	
<p>Standard 3.2.3. Children demonstrate initiative in daily activities (3 indicators)</p>	<p>(Performance Indicators):</p> <ul style="list-style-type: none"> • Undertakes activities in his/her own ways • Displays his/her experiences in various ways (role playing, acting, stories telling, drawing, collaging, movement...) • Suggests new activities
<p>DOMAIN 2 SOCIAL AND EMOTIONAL DEVELOPMENT</p> <p>Sub-domain 2.1. Emotional Development</p>	
<p>Standard 2.1.1: Children are able to perceive themselves (4 indicators)</p>	<p>(Performance Indicators):</p> <ul style="list-style-type: none"> • Child tells important information about his/her self and family members (e.g full name, birthday, address, telephone number, father and mother’s full names, occupation, etc) • Child expresses his or her own preferences for example favourite activities, foods, etc. • Child suggests activities that demonstrate his or her perceived abilities for example playing an outdoor game. • Child differentiates between his or her own preferences and preferences of others e.g. “I like sweet potato, she likes eating corn”.

The Standards approach requires each country to develop its own set of early learning standards that are culturally appropriate. It is far better to develop standards that are appropriate to the national environment than use a measure developed somewhere else that has no

relationship with the country's values for its children. However, experience has shown that it is very helpful for countries to see what others have done and use these standards to help define their own.

The steps for development of Standards are:

- ❖ initial decision making;
- ❖ developing the standards;
- ❖ validation;
- ❖ implementation.

UNICEF's team has been working with over 40 countries to develop standards. Many of them are now being validated in for each age group. This process can take between 3 months and one year, depending on interest and the breadth of the effort. The more ages selected, and the more domains, the longer it will take. The process should be participatory and country specific.

Linking child-level Standards to program standards

Early learning and development standards are used for many purposes:

- ***For individual children's development:*** teacher or health worker assesses what the child can do and decides on a learning plan for the child's development;
- ***For curriculum development:*** used to decide on what kinds of lessons and experiences should be included;
- ***For program quality:*** used for designing teacher training methods, supervision criteria, helping first grades recognize what should be in their curriculum; developing systems for accountability in the program;
- ***For planning:*** determining where resources are most needed (because children are least well prepared according to the criteria), and allocating them there;
- ***For advocacy:*** providing the public with greater understanding of child development and helping them recognize what percent of children might be considered "ready for school;"

- ***For monitoring and program evaluation.*** used to develop a monitoring or assessment system, as was done in Cambodia (Rao et al., 2007).

Using Standards for assessment of learning

If standards are used effectively in classrooms, they assist teachers in focusing on goals for individual children, planning activities to achieve those goals, monitoring the child's progress toward the goal, and assessing the child's progress periodically. This approach to preschool education should result in individualized, age-appropriate and effective learning experiences for children. Given the constraints faced by many programs for disadvantaged children in developing countries, however, only a small portion of these activities may be possible.

In order to use standards for population level or individual level assessment, it is necessary to translate them into an assessment form. They can be collated at the individual item level, to assess learning and progress on each item. Creating a single scale or test from these standards requires a second step of test creation as discussed below.

Pros of Standards approach

- ❖ **Culturally appropriate.** These measures have been defined by each country, and therefore are appropriate for them.
- ❖ **The process increases understanding of early child development.** For countries who have developed these standards, the process of coming up with local standards, and agreeing within a group about what children should know and be able to do before entering school is valuable for planning, program development, and policy development for young children.

Cons of Standards approach

- ❖ **Time-intensive and requires long-term follow-up.** This is a time-intensive process that may take as much as a year, to both develop the standards and complete an age validation (to see if indeed children are able to perform as the standards recommend).
- ❖ **Indicators are not easily translated into a test.** Indicators as developed by a standards-writing team often tend to be too vague to be able to use as a test item. In order to adapt the standards to a test, more work needs to be done to clarify and specify the indicators clearly enough to justify a test.

- ❖ **Needs to be done slowly and carefully.** Good quality information is hard to come by, and the process requires time and care. Someone in a country needs to have this as a priority for it to get done.

CHAPTER 6: TRAINING AND QUALITY CONTROL

Introduction

It is imperative that the research team provide adequate training to testers and supervisors. Trainees should have completed schooling in related disciplines (social sciences, psychology, child development, education) or have relevant experiences (interviewing; community work). It is essential that all testers receive the same training by the psychologists and team on all aspects of the testing situation: approaching families and establishing rapport; introducing the test to families; reading of instructions; administering items and recording responses; offering praise and encouragement; using probes during the administration; and providing feedback on test performance or results.

Connection with local psychologist

As mentioned above local psychologists must be involved with the adaptation and training process. In addition to their necessary inputs during adaptation and training, they will be able to provide continued follow-up training as needed as well as supervision. Universities and local non-government or government agencies are good sources for finding psychology-trained personnel to assist with adaptation and supervision.

Inter-rater reliability

Trainees should also undergo some standardization exercises. For the exercises described below, a “gold standard” interviewer should be established. This person should be trained and efficient with the questionnaire and fluent in the local language. The goals of standardization are to compare each of the trainee interviewers with this gold standard to insure accuracy and reliability. It consists of two parts.

How to test inter-rater reliability

Inter-rater reliability is how much scores among raters agree. This type of reliability is important to ensure that all personnel are administering the assessments in the same way, and

subsequently reduce measurement error or bias due to a particular assessor. To test inter-rater reliability, all interviewers should be present at the same session with the same child or interviewee. The gold standard interviewer (GS) should conduct the assessment with the child or respondent, and record the responses. The trainees, who will follow along silently, also record the responses on their own forms, based on their observations of the assessment (see **Figure 4**). Each trainee’s responses should be compared with those of GS to ensure a correlation of at least 0.80.

To compute the correlation, responses to each item must be compared, and total agreement summed. Let’s say that the GS and trainees assessed a child with a 20 item measure on language development. Each item can be scored as 1 (Pass) or 0 (Fail). Record in a spreadsheet column or on a piece of paper the GS’s responses for item 1, item 2, etc., through item 20. In subsequent columns, record each trainee’s responses for each item. Out of the 20 items, count the number of time in which a trainee’s responses agrees with the GS’s responses. A trainee who has 17 responses in agreement with the GS’s responses would have a correlation of 0.85 (17 divided by 20).

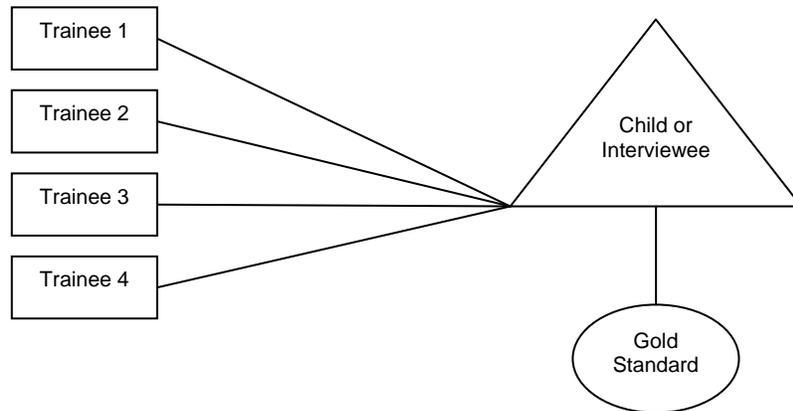


Figure 4: Assessing inter-rater reliability

How to test rater accuracy

In addition to how much trainees agree with one another and the GS, we are also interested in ensuring that each rater is *accurate* in their assessments using a particular measure. To do this, the GS should conduct the assessment or interview with three children or interviewees privately and record the responses to each item (see **Figure 5**). Subsequently, each of the trainees should assess or interview *one of the three* respondents (R1, R2 and R3 in **Figure 5**) individually and

record his/her responses. Each trainee's responses are then compared with those of the GS, and correlations for agreement are computed as described above. A correlation of 0.70 or above is desirable.

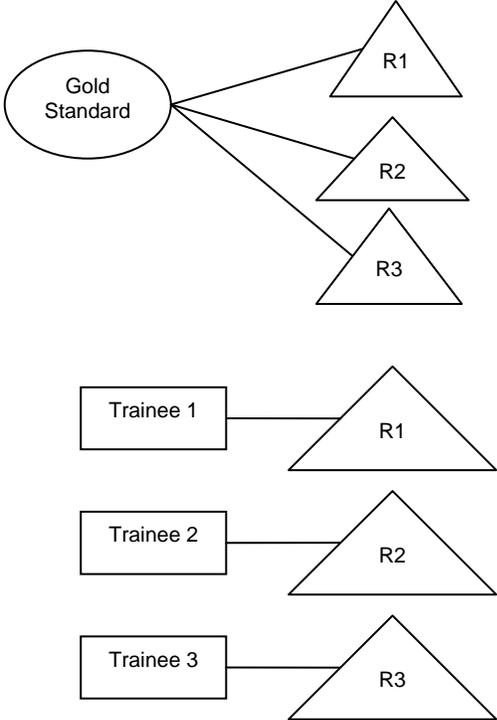


Figure 5: Testing accuracy

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS

Conclusions and future work

Assessments of early childhood cognitive, language, and motor development should be included in the evaluation of education, nutrition or health interventions targeting children under five years old. Although many evaluations to date have focused on physical growth (e.g. height and weight) in children, the development and progression of cognition, language and motor skills are critically important outcomes to measure as well.

Measures of executive function and self-regulatory skills should be included in comprehensive assessments whenever possible because these are likely to provide useful information about children's development and these outcomes are getting increasing attention due to their links with socio-economic status. Researchers should aim to expand our understanding of these domains of child development, and how to measure such sensitive outcomes in difficult field conditions.

Similarly, measures of socio-emotional development should be included in comprehensive assessments whenever possible. Although tests of socio-emotional functioning and development are among the least well-developed of all the developmental domains and can be difficult to adapt cross-culturally, they can offer invaluable insights in the progress of development. Future work should aim to modify and adapt measures of socio-emotional functioning to better incorporate these sensitive and critical outcomes into assessments of nutrition, health or educational program effects.

A key limitation of using tests from developed countries is the inability to use the norms from those countries. In future work, information about normative patterns of development should be collected in children from higher socio-economic groups even if the group of interest is of a lower socio-economic status. If it is possible to collect information from a "norming" sample of children of higher socio-economic status from the country of interest, however, then it may be possible to use that sample as a comparison group representing potential development.

Longitudinal information should be collected when possible and future investigations should aim to collect a data across as long a time horizon as possible. Longitudinal information

is critical for accurately assessing developmental trajectories and change over time can be a much more meaningful outcomes measure than a single point in time.

Broad recommendations

Successful program evaluations (e.g., early childhood education, literacy, or nutrition) hinge on accurately assessing children's development. As mentioned previously, the majority of the assessments reviewed and presented in this toolkit are for child-based measures that occur through an individual (one-on-one) assessment of a child. While we recommend that assessments at the population-level are also necessary and important, there are few population based measures of early childhood development that do not involve an individual assessment of a child. Thus, the majority of the recommendations presented in the toolkit will have to be adapted for use at the population level by examining the data in aggregate.

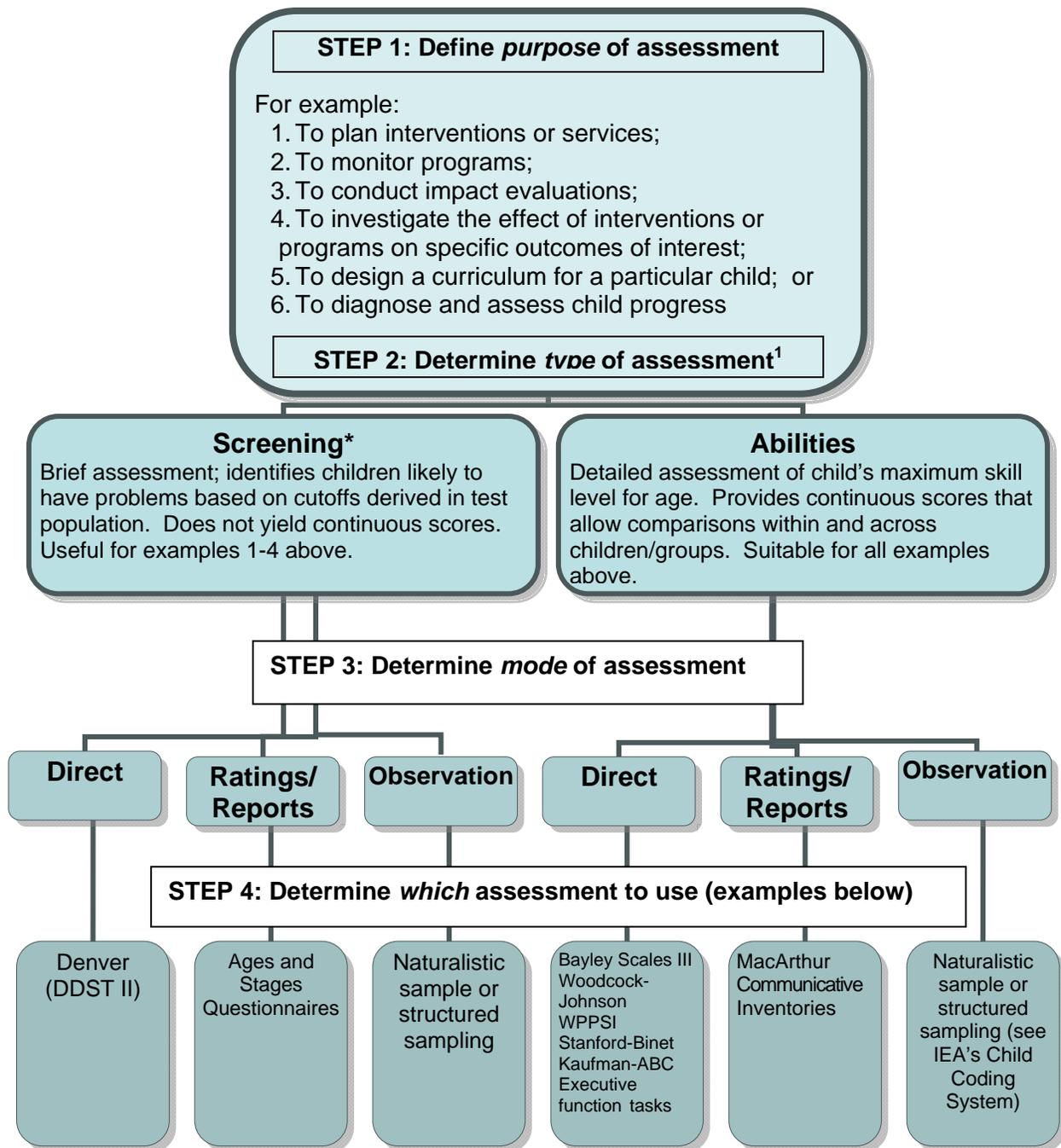
Based on a review of research on young children's development and the results of intervention evaluations, we propose the following broad recommendations, and then follow the broad recommendations with specific recommendations relating to specific test domains and age groups.

- ❖ ***Assess characteristics of the child that the intervention is intending to affect.*** The most important factor to be taken into account in assessment is being sure to measure behaviors that the intervention is hoping to change. For example, an intervention may focus on literacy, and then the appropriate assessment instrument would be a measure of literacy. Similarly, if an intervention is using iron supplementation to help promote cognitive development, then measures of cognition most directly affected by iron status should be used.
- ❖ ***Decide on the type of outcome measure that is appropriate for the evaluation.*** Decide whether the purpose of the assessment is to screen for developmental delay or to have a quantitative measure of development. Decide whether the goal is to have a measure of the population, or an individual-level assessment. Decide whether it is more important to make a comparison within a culture (e.g., comparing an intervention and control group) or a comparison across cultures (e.g., developing a global assessment of children's development).

- ❖ ***Rely upon multiple measures of children’s development.*** In addition to providing a more comprehensive picture of children’s development, some measures index children’s current development (such as the Denver Developmental Test) while others may provide an indication of how children will perform in the future (such as tests of executive function). Some effects of interventions are not apparent until years after the intervention (known as “sleeper effects”). For these reasons, measuring multiple domains of development is especially critical if there are plans to longitudinally examine intervention effects.
- ❖ ***Include assessments of executive function.*** The ability to think, remember information, and engage in the other complex cognitive functions that underlie reading, writing and mathematics are dependent on the development of attention, memory and executive function. Self-regulation encompasses children’s abilities to focus their attention, maintain diligence when faced with difficult tasks, and control their negative emotions when frustrated or angry. These domains are often excluded because there are not many published or standardized tests to assess them. In spite of these limitations, we strongly recommend including these measures given their critical roles as potential mediators of development.
- ❖ ***Consider the cultural context and how it may affect children’s development and school readiness.*** While the tests recommended here have been used in many countries, much less is known about their validity and reliability in developing world contexts. Therefore, it is important for evaluators to have a strong sense of the skills and competencies that are emphasized within each culture, to aid in the interpretation of the data. It is also recommended that evaluators work closely with child psychologists and/or education specialists working in the culture where the evaluation will take place.
- ❖ ***Look for national level tests where possible and use parent/teacher report when possible.*** National level tests can be more appropriate to the context than the adaptations of Western tests described here. A number of these have been included in the Appendices for examples. Assessing children individually with standardized techniques can be time consuming and take a lot of training by skilled professionals. Reports made by teachers, parents or home visitors may be useful as well.

- ❖ *Begin following children early in life.* Results of intervention programs in developing countries indicate that developmental trajectories begin early, and therefore, the greatest insight into the effects of intervention programs may be gleaned from studies that include cohorts of children in infancy and early childhood.

Defining the purpose of the assessment, the type of assessment, the mode of assessment and which actual assessment to use are the key steps that must be completed before assessing a child. (**Figure 6**).



Constraints to consider: budget; copyright issues; time allocated for assessment; training needs and administrator capacities; test setting; capacity of respondents; language and cultural differences requiring extensive adaptation of assessment; materials required for administration. *Screening test cutoffs must be developed within population.

¹Note: Any test could be used as a population measure by aggregating across groups.

Figure 6: Flowchart for decision-making regarding assessment of early childhood development.

Specific recommendations

We have developed an extensive list of recommendations for assessments of child development, both in the realm of published and/or copyrighted tests, and in tests that should be considered in spite of the fact that they have not been developed by testing companies. In **Appendix A** are tests that have been published and/or copyrighted and have been used outside the United States. Appendix A provides an overview of the name of the test, domain of assessment, age range of test, whether the test has been normed, who publishes the test, how to administer and train to administer the test, where it is most often used, time needed for administration of the test, and cost. **Appendices B, C and D** provide an overview of the same information for tests that may not have been published or copyrighted; these tests are included because they are assessments of critical domains (such as executive function) that have often been overlooked by large testing companies. The other advantage to using these tests is often that they are free. The **Table** (available in an Excel spreadsheet) displays all the published studies (obtained searching PubMed, EconLit, PsychInfo, Google Scholar, and Global Health) that have used assessments of early childhood development outside of the US. For over 300 published studies, the table presents: country where the test was used, age range, language of usage, purpose of the study, results from the studies, author and year of the study with a complete reference. In the text below are summary recommendations that highlight the most appropriate tests and more details are provided in the Appendices and in the Table.

Our recommendations required that assessments were:

- ❖ **Psychometrically adequate, valid and reliable;**
- ❖ **Balanced in terms of number of items at the lower end** to avoid children with low scores;
- ❖ **Enjoyable for children** to take (e.g. interactive, colorful materials);
- ❖ **Relatively easy to adapt** to various cultures;
- ❖ **Easy to use in low-resource settings**, e.g. not requiring much material;
- ❖ **Not too difficult to obtain or too expensive;**
- ❖ **Able to be used in a wide age range.**

Infants/Toddlers (Birth to 36 months)

For infants and toddlers, obtaining a *comprehensive* assessment of development is desirable. Due to the interconnections among domains of development in young children, comprehensive assessments provide the greatest sensitivity to intervention efforts, and are also the most time- and cost-effective assessments available. Assessments can be obtained by mother or guardian recall (i.e. asking the mother what the child can or cannot do, or what the child does or does now know), or by direct observation of the child. Although observing the infant or toddler directly is the ideal option to avoid recall bias by the mother or guardian, there are several challenges to this approach. Challenges include greater training necessary for the interviewers in child development and developmental psychology, test/re-test reliability issues (i.e. whether the interviewer could get the same score if s/he administered the test twice), and consistency of timing of the assessment (e.g. between naps for the child, before/after eating, morning/afternoon).

Primary recommendation for individual assessment

Bayley Scales of Infant Development (BSID)

The most commonly used assessment of infant development in the world is the Bayley Scales of Infant Development (BSID). We found 44 published studies that used the BSID outside of the United States, with translations of the test in Spanish, Bahasa, Amharic, Bengali, Japanese, Kiswahili, Chinese, Italian, Turkish and more (see Table). The test has been shown to be sensitive to many types of interventions. Scores on the BSID are normed for children in the United States from 2 to 42 months.

The BSID has been well-validated and provides a strong indication of how children are developing when the test is administered. Performance on the Bayley Scales of Infant Development has been correlated with scores on tests indicative of later academic achievement, including the McCarthy Scales of Children's Abilities, the full scale of the Wechsler Preschool and Primary Scale of Intelligence-Revised, the Differential Ability Scales, and the Preschool Language Scale-Third edition (Bradley-Johnson, 2001). Longitudinal studies have shown that tests similar to the BSID administered as early as 22 months are associated with education outcomes in adulthood, and that the associations between early performance and later outcomes are the strongest in children from households of low socio-economic status (Feinstein, 2003). It

is important to note, however, that scores on the BSID should not be taken as an indication of a child's future IQ, as there is variation in the test's predictive ability.

There are three versions of the BSID (BSID I, II and III). The BSID III was developed very recently to replace the BSID II, which was developed in 1995 to replace the 1969 BSID I. All versions of the BSID provide scores for both the Mental Development Index and the Motor Development Index. The BSID also assesses a child's social and emotional development, by having the tester rate the child's performance during the test (BSID II). The BSID requires a trained assessor, takes about an hour or more to administer, and tends to be expensive. It provides a positive environment, however, with the primary caregiver actively assisting in the testing. Because children's scores on the BSID are not necessarily stable across development, it may be necessary to test children more than once to get a reliable estimate of their developmental status.

The newest version includes language, cognitive, social-emotional, motor and adaptive behavior (caregiver report) subscales that can be scored separately, so that domain-specific assessments can be made. The other major change is that the cognitive subscale requires primarily non-verbal responses from the child, meaning that a child's expressive language skills have less of an impact on their performance on the cognitive items. The third version was standardized with a representative US sample of 1,700 children, and data indicate that the tests measures skills accurately and reliably for children older than six months of age. Testing with very young infants tends to be less reliable and accurate, regardless of the instrument; the lack of reliability in this age range is not surprising (Albers & Grieve, 2007). The inclusion of the separate domain subscales makes the latest version of the Bayley increases the flexibility of how the scales can be used (i.e., one can pick and choose which subscales to administer). An additional benefit to this is that the amount of time required for administration is shortened (compared to earlier versions) if one chooses to not administer all five scales. While there is not yet a lot of information on its use in the US or in developing countries, it is anticipated that the latest version will continue to be sensitive to the effects of interventions.

Alternative measures for individual assessment

The MacArthur Communicative Development Inventories (CDI)

CDIs are parent report forms for assessing language and communication skills in infants and young children. This scale has been shown to provide valid assessments of early language milestones in young Spanish-speaking children,(Marchman & Martine-Sussmann, 2002) and has been linked with important biological outcomes. It was adapted for Bangladesh by Hamadani et al. (Hamadani, et al., Submitted).

Kilifi Executive Function Scale (Kilifi Developmental Inventory)

Working memory and inhibition are two types of executive behaviors that are measurable in infants. The Kilifi executive function scale includes the A not B task (described above in the Executive Function section) and the Self-control task. The A no B task measures both working memory and inhibition. In this task, the infant watches as a toy is hidden in one of two locations; after a brief delay, the infant is encouraged to retrieve the hidden toy. After a number of successful trials, the toy is hidden in the alternate location. The Self- Control task assesses inhibition. For this task, the child is shown a gift box. The assessor then tells the child not to take it until instructed to do so (which occurs after a pre-determined amount of time passes). How long the child is able to delay reaching for the gift is recorded, as well as whether the child waited the entire interval.

Alternative measures for screening purposes

Ages and Stages Questionnaires (ASQ)

The ASQ, often used in the United States by home visitors to determine delay or recommend intervention, is a low-cost and easily administered, comprehensive checklist of developmental milestones. The ASQ is parent report, and can be completed by parents alone or administered by a trained assessor. The subscales measure skills in Communication, Gross Motor, Fine Motor, Personal-Social and Problem-Solving (similar to cognitive) domains. The questionnaires are divided into two- to three-month age intervals for use with children 4-60 months of age. Scores are normed to indicate whether children are developing age-appropriately, but it does not provide standardized scores as are available for the BSID. It is both less detailed and less validated than the BSID, but also may offer an opportunity to systematically obtain information about when children are reaching developmental milestones in diverse contexts. There are published reports of its use in Ecuador, and we are aware of its use in several other countries (unpublished).

Denver Developmental Screening Test (DDST)

The DDST is a comprehensive test of children's development, and can be used to assess development from birth through 5 years of age. We found 15 published papers that have used it in the developing world in countries as diverse as Armenia, Brazil, China, Turkey and Zaire. It requires a trained administrator, but in contrast to the ASQ, it has been used extensively within the developing world. Scores from the DDST indicate how children are developing in four domains: Fine motor/adaptive, gross motor, language, and personal/social. The DDST does not provide continuous scores indicating children's developmental status, instead only providing an indication of whether the child appears to have developmental delays when compared to children of the same age. Thus, investigators should be cautioned against using the Denver out of the context of screening.

Evaluacion de Escala de Evaluacion del Desarrollo Psicomotor (EEDP)

The EEDP is a Spanish-language screening test initially developed in Chile and widely used in Latin America. It has assessments in the areas of language, social, coordination, and gross motor. Children are divided into three categories: normal, risk, and delayed.

The Guide for Monitoring Child Development

This parent report assessment provides a method for developmental monitoring and early detection of developmental difficulties in children of low and middle income countries. The questions are designed to be simple and clear. The caregiver is respondent, who completes a brief, open-ended, pre-coded interview. The questions pertain to child's social, emotional, and cognitive development.

Summary: Recommended tests for children 0-2

Continuous measure, direct assessment

	<i>PROS</i>	<i>CONS</i>
Bayley Scales of Infant Development III	<ul style="list-style-type: none"> -Comprehensive assessment -Separate subscales increases flexibility of use -Inclusion of parent report adaptive scale allows parent involvement in the assessment process -“Gold standard” for infants 	<ul style="list-style-type: none"> -Expensive (\$1000 for initial kit and time intensive for interviewers) -Needs large amount of equipment -Requires training of interviewer -60-90 minutes required for administration (when all 5 subscales are administered) -Materials and words need to be adapted with care to the setting.
Nationally adapted test (e.g. Indian version of Bayley II) Philippines ECCD Checklist	<ul style="list-style-type: none"> -Already culturally adapted 	<ul style="list-style-type: none"> -National adaptations may use older versions of test (e.g. Bayley I, which is out of date).
Kilifi Executive Function Tasks	<ul style="list-style-type: none"> -Fairly quick and easy to use -Easily adaptable to various contexts -Minimal materials needed -Free 	<ul style="list-style-type: none"> -Training required

Continuous measure, maternal report

	<i>PROS</i>	<i>CONS</i>
MacArthur Communicative Development Inventories	<ul style="list-style-type: none"> -Easy and quick to administer (to mother) -Free 	<ul style="list-style-type: none"> -Only measures language development -Administered to mother, which may result in recall bias -Has not been used widely in languages other than English and Spanish
Guide for Monitoring Child Development (Turkey)	<ul style="list-style-type: none"> - Parents’ report to physicians about concerns about child; easy to administer - designed to create partnership of parent and physician -appropriate to lower-middle income countries 	<ul style="list-style-type: none"> - Appropriate to lower-middle-income countries; may not work in very poor countries

Screening test, direct assessment		
	<i>PROS</i>	<i>CONS</i>
Denver Developmental Screening Test	<ul style="list-style-type: none"> -Assessment of some domains of development -Cheaper than Bayley (\$90 for initial kit) -Has been used widely in the developing world -Appropriate for children up to 5 years of age 	<ul style="list-style-type: none"> -Not designed to assess specifics of any particular construct (e.g. language) -Requires trained administrator -Does not yield continuous scores
Nationally developed test (e.g. EEDP, Chile; Schoklo, Thailand)	<ul style="list-style-type: none"> -culturally appropriate items -able to administer it in low resource, briefly trained testers 	<ul style="list-style-type: none"> - may only be appropriate to the particular country
Screening test, maternal report		
	<i>PROS</i>	<i>CONS</i>
Ages and Stages Questionnaires	<ul style="list-style-type: none"> -Most items are easily modifiable for cultural context -Language of items is fairly simple (5th grade level) -No concern about timing because administered to mother -Cheaper than Bayley (\$200 for initial kit) -Easy 	<ul style="list-style-type: none"> -As with all maternal reports, bias is possible -Some items may not be easily observed by mothers in developing countries (e.g., child's response when looking in mirror) -Reports on its use in the developing world are scarce to date, but it is currently being used in several countries
Guide for Monitoring Child Development	<ul style="list-style-type: none"> - can also be used as a screening test (see above) 	<ul style="list-style-type: none"> - Does not include a wide range of outcomes

Preschool-aged Children (3 to 5 years)

As children grow older, it becomes more critical to assess language and cognitive development, as well as children's abilities to behave appropriately and regulate their attention and emotions. Accordingly, our recommendations include a *wider range* of assessments than we recommend for younger children. Another difference is that the assessments are broken down by domain.

Cognitive or comprehensive assessments

Tests of cognitive ability in young children can focus on aptitude (IQ) or achievement (knowledge related to school readiness, such as letter or word identification), or a combination of these factors. Tests that focus more heavily on achievement, or knowledge, are arguably more sensitive to environmental effects such as exposure to high-quality parent-child interactions and language use within the home. Results from intervention programs in the United States have shown that children's knowledge, as indexed by achievement tests, is enhanced by early intervention. Conversely, IQ is thought to be less affected by children's social or familial environments, but is perhaps more significantly affected by children's neurological functioning. Since IQ measures mental age and mental age is influenced by both neurological maturation and environmental input, it is definitely influenced by early interventions -- especially in countries where the normal input is low. The effects on IQ are less than on achievement tests because the latter include skills that are intentionally taught in a preschool whereas cognitive developmental tests assess skills learned incidentally

Therefore, tests that place at least some emphasis on both achievement and aptitude are perhaps more likely to reflect the results of an intervention. Because they are responsive to environmental influences, however, it is important to note that tests of achievement may be more culture-bound, and therefore less appropriate to use across diverse contexts. In addition, children's scores on IQ tests tend to be somewhat unstable before age six, which reduces their usefulness for measuring intervention effects (Brody, 1992).

British Abilities Scales II (BAS II)

The BAS II was launched in 1996 as a further development of the initial British Ability Scales in 1979 and the 1990 Differential Ability Scales, the US version. The purpose was to be able to assess a variety of abilities rather than a single measure of intelligence to improve the

capacity of the test to assist with information on functioning through the age range. The scales are derived from the information-processing model of Horn-Cattell (Elliott, 1996; Hill, 2005) which suggests that there are many different kinds of abilities including a core element of general fluid intelligence or “g”. The Early Years battery ranges from age 2.6 through 7 years, and the School Age scales go through age 17.2. At each age, there are core tests measuring “g”, diagnostic tests measuring specific skills like memory or visual recognition, and at school age, tests of achievement. Tests can be combined to create a cluster, or specific subtests can be used. Considerable efforts have been taken to ensure that the tests are appropriate for diverse social, racial, and linguistic backgrounds. The standardization sample is representative of the UK population. It requires relatively little verbal expression of the child, and covers an age range of 2 years through elementary school age, making it very useful for a wide range of testing. It has been used to evaluate the Madrasa Resource Center preschools (Aga Khan Foundation) in Uganda, Kenya, and Zanzibar (Mwaura, Sylva, & Malmberg, 2008) and the Nutrition and Early Childhood Development program in Uganda. It has also been used in India and Zimbabwe (Mpofu, 1995).

Grover-Counter Scale of Cognitive Development-Revised (South Africa)

The test measures level of cognitive functioning (within defined range) of persons with impaired verbal skills, whether receptive, expressive, or both. It helps with diagnosis and treatment for mentally handicapped. The final norms were established predominantly from data derived from 1) 200 children from three to 10 years of age, the majority of whom were White, and 2) 419 children from four different South African provinces.

ICMR Psychosocial Developmental Screening Test (India)

Based on passing rates of milestones from 10,000 children in India, this test screens children for delays in five major developmental areas: 1) gross motor, 2) vision and fine motor, 3) hearing, language and concept development, 4) personal skills, and 5) social skills. One can also obtain a continuous score.

Kaufman ABC (K-ABC)

K-ABC is an intelligence test of problem-solving ability which is normed for children’s performance on three subscales: achievement, simultaneous processing (ability to solve problems by integrating diverse pieces of information simultaneously), and sequential processing

(ability to solve problems by ordering items or placing them in sequence). The K-ABC has been used in a handful of studies evaluating the effects of intervention programs, and has shown sensitivity to changes in nutritional status, including iron and iodine; the K-ABC has also shown sensitivity to exposure to malaria. It has been used in several different languages, including French (in Benin), Laotian, Wolof (spoken in Senegal) and Kikongo (spoken in Zaire).

Leiter International Performance Scale

The Leiter is also a test of intelligence; it is generally not considered a test of achievement or knowledge. The Leiter emphasizes fluid intelligence and is non-verbal, and therefore may be easier to use and interpret across diverse contexts. To date, however, there have been few studies using the Leiter, outside of the United States; it has only been used in Saudi Arabia, Taiwan, Italy and Spain.

McCarthy Scales of Children's Abilities (MSCA)

MCSA is a comprehensive battery that offers a broad picture of a child's abilities with attractive materials and carefully designed game-like tasks suitable for children of both sexes and from various ethnic, regional and socio-economic backgrounds (Boivin, et al., 1995). The gross motor sub-scales of the McCarthy Scales are well-suited for assessments of toddlers and pre-school children. For toddlers, gross motor skills include learning to walk and run, and for preschool-aged children, gross motor skills include walking on a line, controlling movements in games, hopping on one foot and jumping. Although the timing of most large motor skills is not indicative of future development, a failure to demonstrate these skills may indicate the presence of a developmental delay. The MSCA has been used in Mexico, Jamaica, France and the Seychelles.

Philippines Early Childhood Care and Development (ECCD) Checklist

The Philippines ECCD Checklist was developed within country after an extensive process of piloting and the development of norms. The test monitors child's development in the following domains: fine and gross motor, receptive and expressive language, self-help, cognitive, and social-emotional. It was normed on a sample of over 10,000 Filipino children. It is a continuous measure.

Stanford Binet

The Stanford-Binet is a test of intelligence and is generally not considered a test of achievement or knowledge. The Stanford-Binet measures fluid reasoning, knowledge, quantitative reasoning, visual-spatial processing, and working memory. The Stanford-Binet does not appear to have been used extensively in studies of interventions, and so it is not possible to determine how sensitive it may be to intervention effects. It has been used in a couple of studies in India, however, and three others in Asia (Japan, Thailand and China or Taiwan). It has also been used successfully in Madagascar (unpublished).

Test de Desarrollo Psicomotor (TEPSI) (Chile)

The TEPSI evaluates development in three basic areas--motor function, coordination, and language--by observing child's behavior in certain situations set up by the examiner. The test is continuous, but may be used as a screening measure. (There are standards for "normal," "at risk," and "delayed" children.) The child is asked to perform various activities based on the area of development in question. For motor function, for example, the child is asked to perform either a prolonged action or sequence of actions.

Weschler Preschool and Primary Scale of Intelligence (WPPSI)

The WPPSI is an extension of the Weschler Intelligence Scale for Children. Both are designed to be measures of intelligence, not achievement. There are two broad factors on this scale: performance and verbal. Performance items do not require that the child talk to the experimenter and so may be less sensitive to cultural biases and easier to use across diverse linguistic contexts. The WPPSI has been used widely around the world, including Brazil, China, Iran, Mexico, Pakistan, and Venezuela.

Woodcock-Johnson (WJ)

The WJ is a normed set of tests for measuring general intellectual ability, specific cognitive abilities, and scholastic achievement. The scales have previously been translated into Spanish and adapted for Latin American contexts and have been used to evaluate effects of early childhood nutritional interventions and early health insults on cognitive development in infants and older children. The Woodcock-Johnson tests have shown sensitivity to an early intervention program aimed at low income families, and can pick up differences between children who were born low birth weight when compared with normal weight children. Many other investigators have documented the changes in scores on the Woodcock-Johnson tests to interventions, such as

changing eating patterns at home, including the increased intake of milk and other animal products. Not many investigators have used the Woodcock-Johnson outside of the context of the United States, however, and it appears only to have been translated into Spanish and a French-based Creole.

Language only

Peabody Picture Vocabulary Test (PPVT) / Test de Imágenes de Vocabulario Peabody (TVIP, in Spanish)

The PPVT (is a test of “receptive language” or listening comprehension for the spoken world and has been used in many countries throughout the world, including China, France, Jamaica, South Africa and the Seychelles to name a few. In the test, the child is shown four pictures (e.g. dog, fork, doll and table) and is asked to point at one of them. The test has been translated and normed in Spanish; items have been carefully selected through rigorous item analysis for their universality and appropriateness to Spanish-speaking communities. Children’s scores on the PPVT have been shown to be sensitive to early intervention efforts (Love, et al., 2005). The TVIP is frequently used to evaluate the language development of Spanish-speaking preschool children, and older students. The Peabody Picture Vocabulary Test has also been adapted and used in a four-country longitudinal study, “Young Lives”, in Peru, Vietnam, India, and Ethiopia. (See <http://www.younglives.org.uk/publications/technical-notes>, No 15.) Care should be taken to make sure that the pictures are comprehensible, especially among children who have not been exposed to pictures.

Reynell Developmental Language Scale

The 134-item Reynell scale (Reynell, 1990) is comprised of two subscales to assess both Receptive Language and Expressive Language. It is administered individually with the child and uses picture, toys and puppets to elicit responses. The Receptive Language subscale measures how a child responds to verbal requests to perform an activity, such as “Put the doll on the chair.” The Expressive Language subscale assesses three aspects of children’s speech: structure (e.g., use of pronouns, past tense), content (e.g., use of language to describe a picture) and vocabulary. The Reynell scales have demonstrated excellent reliability in a large sample of low-income children (McCartney, Dearing, Taylor, & Bub, 2007), and have also been shown to be predictive of intelligence scores in the UK (Silva, 1986). It has been used clinically in South

Africa, and is currently in use in a large epidemiological study (Jane Kvalsvig, personal communication).

Motor skills

Fine motor skills: e.g. pegboard

Fine motor skills include such abilities as picking up objects and holding eating utensils, threading a bead, drawing a circle. For preschool-aged children, fine motor skills include the ability to hold a pencil, write and draw. Thus, it is relatively easy to design a test for fine motor skills. The pegboard test is one type of test that assesses a child's ability to exercise hand-eye coordination and fine motor control, and requires that children pick up pegs and place them into a board with peg-sized holes. For more advanced or older children, the pegs can be made as keys with a correct and incorrect orientation. Thus, the child must first orient the peg and then insert it into the hole.

WHO Motor Milestones Assessment.

For children 0-2, this assessment measures 6 milestones with a carefully defined protocol for testing and recording. Data are available on well-nourished children from 6 countries.

Executive function

Leiter Examiner Scale

Some aspects of executive function processes can be assessed by having test administrators rate children's performance during the assessment, on such items as children's negative affect, attention to tasks, and orientation to the examiner. These ratings are supplemented by questions to parents about children's typical behavior. The Leiter Examiner Scale has been used among low-income children in the United States (for the evaluation of the Head Start program). While it is likely that most children in the developing world have less experience with testing materials and testing situations than children in the developed world, ratings of their behavior during testing situations still may provide insight into their executive function abilities.

Day/Night Stroop test

The Day/Night task is a form of the adult Stroop tests and primarily assesses inhibition and working memory. The assessor talks with children about when the sun rises (in the day) and

when the moon and stars come out (in the night). The child is then shown two cards: a white card with a yellow sun and a black card with a white moon and stars. Children are told that this is a game where they must say “night” when shown the sun card and “day” when shown the moon/stars card. Following some practice trials, there are 16 test trials. Each card is presented in a fixed, pre-determined order. The number of correct responses is recorded (Gerstadt, Hong, & Diamond, 1994).

Backward Digit Span

This test measures both inhibition and working memory. The assessor instructs the child to repeat whatever she says backwards. After a demonstration and practice trial, the assessor suggests they do some more. The trials begin with two digits and increase in the number of digits until children make mistakes on three consecutive trials. The highest level of successful completion is recorded (two, three, four, or five digits) (Davis & Pratt, 1996).

The BRIEF-P

This 63-item rating scale can be completed by parents or teachers. The items cover five executive function skills: Inhibition, Attentional Shift, Emotional Control, Working Memory, and Planing/ Organization. Three index scores that summarize the scales are provided for Inhibitory Self-Control (Inhibit and Emotional Control), Flexibility (Shift and Emotional Control), and Emergent Metacognition (Working Memory and Plan/Organize). The BRIEF-P has a reading level at approximately the fifth-grade and takes 10–15 minutes to complete. Respondents’ rate each item as to whether is never, sometimes, or often a problem for the child. Norms by age and sex are available, and standardized scores and percentiles are available. The scale has shown good psychometric properties (Gioia, Isquith, Retzlaff, & Espy 2002).

Social and behavioral development

Strengths and Difficulties Questionnaire.

This questionnaire has been translated into several languages, and has been shown to be reliable across parental education levels in identifying children with clinically-relevant mental health problems. The SDQ may be preferable over the Child Behavior Checklist because it has already been translated into several languages and has been shown to be reliable in diverse populations. Furthermore, it is freely available (see <http://www.sdqinfo.com>). The SDQ has

been used in Brazil, Pakistan, Bangladesh, Israel, Yemen, Thailand, and in the Democratic Republic of Congo.

Achenbach Child Behavior Checklist (CBCL)

The CBCL is a test in which the parent or guardian rates a child's problem behaviors and competencies, and has been used in many countries outside of the US, including Mexico, Turkey, India, Ethiopia and Thailand. It is designed to assess in a standardized format the behavioral problems and social competencies of children as reported by parents and includes questions relating to aggression, hyperactivity, bullying, conduct problems, defiance, and violence at home and at school, and has been used in low-income Spanish-speaking populations.

Early Development Inventory.

This teacher rating form has been used in several different countries and shows adequate internal consistency among countries. The teacher rates all or a sample of children on five different dimensions, including social and emotional development. Children are defined in terms of vulnerability and a cut-off point for vulnerability is defined. The scale allows one to calculate the percent of children in a particular group who are vulnerable.

Summary: Recommended tests for pre-school children

Cognitive development - Recommended		
	PROS	CONS
Stanford Binet	<ul style="list-style-type: none"> -Tests multiple domains of intelligence and a wide range of ages; -Can be used with children as young as 2 years old; -Test is fun and engaging for children; -Has a large enough number of items at the lower end, which means we do not have a “floor”; -Has shown sensitivity to nutrition interventions (e.g. iodine supplementation) - Has been used in many developing countries; -Easy to adapt cross-culturally; -Non-verbal subtests avoid issues of translation 	<ul style="list-style-type: none"> - Some subtests have to be adapted to local context -No information on achievement -IQ scores become more stable as children grow older; may be unstable in young children -Expensive -Requires extensive training -Lengthy administration time
British Ability Scales II Early Years	<ul style="list-style-type: none"> -Provides both general ability scores and a series of subtests; -Test is fun and engaging for children; -Tests a wide range of ages -Does not require much expressive verbalization -Efforts have been made to reduce bias -Specific subtests can be used 	<ul style="list-style-type: none"> -Expensive - Some subtests have to be adapted to local context - Does not measure very young children -Stop/start places and norms are appropriate for UK children but may not be appropriate in other settings.
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	<ul style="list-style-type: none"> -Non-verbal subtests avoid issues of translation - May be less culturally bound than achievement measures -Relatively short administration time -Has been used widely in the 	<ul style="list-style-type: none"> -No information on achievement -Unclear whether IQ scores are sensitive to intervention effects -IQ scores become more stable as children grow older; may be unstable in young children

developing world.

- Verbal subtests require adaptation
 - Expensive
-

Cognitive development - Alternatives

	<i>PROS</i>	<i>CONS</i>
Kaufman ABC	<ul style="list-style-type: none">- May be less culturally biased than other measures because of non-verbal components.	<ul style="list-style-type: none">-Not recommended for use as the primary instrument for identifying the intellectual abilities of children either in research or in clinical settings-Expensive & requires extensive training-Lengthy administration time
Woodcock-Johnson test	<ul style="list-style-type: none">-Comprehensive assessment of aptitude and achievement-Shown sensitivity to many types of interventions-Can be used across wide age range	<ul style="list-style-type: none">-Expensive (\$1000 for initial kit)-Needs large amount of equipment-Requires training of interviewer-Long time for administration
Leiter	<ul style="list-style-type: none">-Non-verbal, avoids issues of translation- May be less culturally sensitive than achievement measures-Includes socio-emotional and executive function assessment by interviewer	<ul style="list-style-type: none">- Not recommended for use as the primary instrument for identifying the intellectual abilities of normal or special children either in research or in clinical settings-IQ scores become more stable as children grow older; may be unstable in young children-Expensive-Requires extensive training-Lengthy administration time
EDI	<ul style="list-style-type: none">- provides a teacher rating scale of vulnerability in 5 dimensions	<ul style="list-style-type: none">- Does not provide individual level information- Teacher ratings may vary by social context
Locally developed tests (ICMR, ECCD, TEPSI, Grover)	<ul style="list-style-type: none">- Depends on the test; may be more appropriate to local context	<ul style="list-style-type: none">- May not have norms that are appropriate to all contexts
McCarthy Scales	<ul style="list-style-type: none">-Assessment of many developmental	<ul style="list-style-type: none">-Not designed to assess specifics of

domains	any particular construct (e.g. language) except for gross motor.
-Cheaper than Bayley (\$90 for initial kit)	-Requires trained administrator
-Has been used widely in the developing world	

Language development

	<i>PROS</i>	<i>CONS</i>
Peabody Picture Vocabulary Test	<ul style="list-style-type: none"> -Has been used widely in the developing world -Easy to administer -Picture-based so no need for extensive translation -Sensitive to a wide variety of interventions 	<ul style="list-style-type: none"> -Some words or concepts may not be culturally appropriate -Children without experience in decoding pictures will tend to score lower. -Requires substantial training for scoring, which must occur in the field -Moderately expensive (\$379 for complete kit)
Reynell Developmental Language Scale	<ul style="list-style-type: none"> -Comprehensive assessment of both expressive and receptive language skills -Subscales can be used alone or together -Attractive, fun materials used to elicit responses 	<ul style="list-style-type: none"> -Requires considerable adaptation -Expensive (~\$500.00)

Motor skills

	<i>PROS</i>	<i>CONS</i>
Pegboard (fine motor skills)	<ul style="list-style-type: none"> -Easy to administer -Easy to design in-country and can be manufactured locally 	<ul style="list-style-type: none"> -Some costs in development of pegboard

Executive function

	<i>PROS</i>	<i>CONS</i>
Leiter Examiner Scale	<ul style="list-style-type: none"> -Easy to include in assessment because scored by interviewer immediately following test (the items on the Leiter could be modified to use with other tests of cognitive 	<ul style="list-style-type: none"> -Requires the administration of the entire Leiter test simultaneously

	ability)	
	-Easily adapted cross-culturally	
Day/Night Task & Backward Digit Task	-Fairly quick and easy to use -Fairly easy to adapt cross-culturally -Measures both working memory and inhibition -Minimal materials needed -Free	-Requires training for use -Has not yet been widely used in developing countries with this age range
BRIEF-P (Parent and teacher report)	-Can be used with parents, teachers, other caregivers -Easy and quick to administer -Does not require materials -Comprehensive assessment of executive function -Relatively inexpensive (\$139.00 per kit)	-Risk of potential bias -No information on its use in developing countries or adaptation

Social and behavioral development

	<i>PROS</i>	<i>CONS</i>
Strengths and Difficulties	-Widely used and translated; easy and quick to administer -Free	-May be cultural modifications necessary – depends on context
Achenbach Child Behavior Checklist	-Administered to the mother -Has been widely used and translated -Well validated	-Expensive, especially if want to modify or translate -Very negatively valenced, i.e. lots of questions about negative and anti-social behavior

REFERENCES

- Aboud, F. E. (2007). Evaluation of an early childhood parenting programme in rural Bangladesh. *Journal of Health, Population, and Nutrition, 25*(1), 3-13.
- Aboud, F. E., & Alemu, T. (1995). Nutrition, maternal responsiveness, and mental development of Ethiopian children. *Social Science & Medicine, 41*, 725-732.
- Abubakar, A., Holding, P., van Baar, A., Newton, C., & van de Vijver, F. (2008). Monitoring psychomotor development in a resource-limited setting: an evaluation of the Kilifi Developmental Inventory. *Annals of Tropical Paediatrics, 28*(3), 217-226.
- Abubakar, A., van de Vijver, F., Mithwani, S., Obiero, E., Lewa, N., Kenga, S., et al. (2007). Assessing developmental outcomes in children in Kilifi, Kenya, following prophylaxis for seizures in cerebral malaria. *Journal of Health Psychology, 12*(3), 417 - 430.
- Abubakar, A., Van de Vijver, F., Van Baar, A., Mbonani, L., Kalu, R., Newton, C., et al. (2008). Socioeconomic status, anthropometric status, and psychomotor development of Kenyan children from resource-limited settings: a path-analytic study. *Early Human Development, 84*(9), 613-621.
- Adolph, K. E. (2002). Babies' steps make giant strides toward a science of development. *Infant Behavior & Development, 25*, 86-90.
- Adolph, K. E., Vereijken, B., & Denny, M. A. (1998). Learning to crawl. *Child Development, 69*, 1299-1312.
- Adolph, K. E., Vereijken, B., & Shrout, P. E. (2003). What changes in infant walking and why. *Child Development, 74*, 475-497.
- Agarwal, D. K., Upadhyay, S. K., Tripathi, A. M., & Agarwal, K. N. (1987). *Nutritional status, physical work capacity and mental function in school children* (No. 6). New Delhi: Nutrition Foundation of India.
- Ainsworth, M. (1993). Attachment as related to mother-infant interaction. *Advances in Infancy Research, 8*, 1-50.
- Alaimo, K., Olson, C. M., & Frongillo, E. A. (1999). The importance of cognitive testing for survey items: An example from food security questionnaires. *Journal of Nutrition Education 31*, 269-275.

- Albers, C. A., & Grieve, A. J. (2007). Test Review: Bayley, N. (2006). Bayley Scales of Infant and Toddler Development- Third Edition. San Antonio, TX: Harcourt Assessment.
- Journal of Psychoeducational Assessment*, 25(2), 180-190.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7 ed.). Upper Saddle River, NJ: Prentice Hall.
- Anderson, V. (1998). Assessing executive functions in children: Biological, psychological, and developmental considerations. *Neuropsychological Rehabilitation*, 8(3), 319-349.
- Atwine, B., Cantor-Graae, E., & Bajunirwe, F. (2005). Psychological distress among AIDS orphans in rural Uganda. *Social Science & Medicine*, 61(3), 555-564.
- Baddeley, A., Meeks Gardner, J., & Grantham-McGregor, S. (1995). Cross-cultural cognition: Developing tests for developing countries. *Applied Cognitive Psychology*, 9, S173-S195.
- Bagnato, S. J., Smith-Jones, J., McComb, G., & Cook-Kilroy, J. (2002). *Quality early learning - Key to school success: A first-phase 3-year program evaluation research report for Pittsburgh's Early Childhood Initiative (ECI)*. . Pittsburgh, PA: SPECS Program Evaluation Research Team.
- Bayley, N. (1969). *Manual for the Bayley Scales of Infant Development*. New York: The Psychological Corporation.
- Black, M. M., Hess, C. R., & Berenson-Howard, J. (2000). Toddlers from low-income families have below normal mental, motor, and behavior scores on the revised Bayley scales. *Journal of Applied Developmental Psychology*, 21(6), 655-666.
- Bloom, L. (1998). Language acquisition and its developmental context. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of Child Psychology, 5th edition. Volume 2: Cognition, Perception and Language*. (pp. 1-50). New York: John Wiley.
- Bogin, B., & MacVean, R. B. (1983). The relationship of socioeconomic status and sex to body size, skeletal maturation, and cognitive status of Guatemala City schoolchildren. *Child Dev*, 51, 115-128.
- Boissiere, M., Knight, J. B., & Sabot, R. (1985). Earnings, schooling, ability and cognitive skills. *American Economics Review*, 75, 1016-1030.
- Boivin, M. J., & Giordani, B. (1993). Improvements in cognitive performance for schoolchildren in Zaire, Africa, following an iron supplement and treatment for intestinal parasites. *J Pediatr.Psychol*, 18(2), 249-264.

- Boivin, M. J., Green, S. D., Davies, A. G., Giordani, B., Mokili, J. K., & Cutting, W. A. (1995). A preliminary evaluation of the cognitive and motor effects of pediatric HIV infection in Zairian children. *Health Psychol.*, *Jan*;14(1), 13-21.
- Bolig, E. E., Borkowski, J., & Brandenberger, J. (1999). Poverty and health across the life span. In T. L. Whitman & T. V. Merluzzi (Eds.), *Life span perspectives on health and illness* (pp. 67-84). Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc., Publishers.
- Bracken, B. A. (2007). Creating the optimal preschool testing situation. In B. A. Bracken & R. Nagle (Eds.), *Psychoeducational Assessment of Preschool Children* (4th ed., pp. 137-154). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bracken, B. A., & Barona, A. (1991). State of the art procedures for translating, validating and using psycho-educational tests in cross-cultural assessment. *School Psychology International*, *12*, 119-132.
- Bradley-Johnson, S. (2001). Cognitive assessment for the youngest children: A critical review of tests. *Journal of Psychoeducational Assessment*, *19*, 19-44.
- Bradley-Johnson, S., & Johnson, C. M. (2007). Infant and toddler cognitive assessment. In B. A. Bracken & R. Nagle (Eds.), *Psychoeducational Assessment of Preschool Children* (4 ed., pp. 325-358). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*, 371-399.
- Bradley, R. H., & Corwyn, R. F. (2005). Caring for children around the world: A view from HOME. *International Journal of Behavioral Development*, *29*(6), 468-478.
- Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & Garcia Coll, C. (2001). The home environments of children in the United States Part I: Variations by age, ethnicity, and poverty status. *Child Development*, *72*(6), 1844-1867.
- Breitmayer, B. J., & Ramey, C. T. (1986). Biological nonoptimality and quality of postnatal environment as codeterminants of intellectual development. *Child Development*, *57*(5), 1151-1165.
- Bretherton, I., Bates, E., Benigni, L., Camaioni, L., & Volterra, V. (1979). Relationships between cognition, communication, and quality of attachment. In E. Bates (Ed.), *The emergence of symbols: Cognition and communication in infancy*.

- Bricker, D., & Squires, J. (1999). *Ages and Stages Questionnaires: A Parent Completed, Child Monitoring System, 2nd Ed.* Baltimore, MD: Paul Brookes.
- Brinkman, S., Silburn, S., Lawrence, D., Goldfield, S., Sayers, M., & Oberklaid, F. (2007). Investigating the validity of the Australian Early Development Index. *Early Education and Development, 18*(3), 427-451.
- Brody, N. (1992). *Intelligence* (2nd Ed ed.). San Diego: Academic Press.
- Brooks-Gunn, J., Klebanov, P., Liaw, F. r., & Duncan, G. J. (1995). Toward an understanding of the effects of poverty upon children. In H. E. Fitzgerald & B. M. Lester (Eds.), *Children of poverty: Research, health, and policy issues* (pp. 3-41). New York, NY, USA: Garland Publishing, Inc.
- Brooks-Gunn, J., Leventhal, T., & Duncan, G. J. (2000). Why poverty matters for young children: Implications for policy. In J. D. Osofsky & H. E. Fitzgerald (Eds.), *Parenting and Child Care* (Vol. 3, pp. 89-131). New York, NY: John Wiley & Sons, Inc.
- Bushnell, E. W., & Boudreau, J. P. (1993). Motor development and the mind: The potential role of motor abilities as a determinant of aspects of perceptual development. . *Child Development, 64*, 1005-1021.
- Carlson, S. M. (2005). Developmentally Sensitive Measures of Executive Function in Preschool Children. *Developmental Neuropsychology, 28*(2), 595 - 616.
- Carter, J. A., Lees, J. A., Murira, G. M., Gona, J., Neville, B. G. R., & Newton, C. R. J. C. (2005). Issues in the development of cross-cultural assessments of speech and language for children. *International Journal of Language & Communication Disorders, 40*(4), 385 - 401.
- Chorover, S. L. (1979). *From genius to genocide: The meaning of human nature and the power of behaviour control.* . Cambridge, MA: The Massachusetts Institute of Technology Press.
- Chun, F. Y. (1971). Nutrition and education - a study. *Journal of the Singapore Pediatric Society, 13*(2), 91-96.
- Clarke, N., Grantham-McGregor, S. M., & Powell, C. (1991). Nutrition and health predictors of school failure in Jamaican children. *Ecology of Food and Nutrition, 26*, 1-11.

- Cohen, D. A., Mason, K., Bedimo, A., Scribner, R., Basolo, V., & Farley, T. A. (2003). Neighborhood physical conditions and health. *American Journal of Public Health, 93*(3), 467-471.
- Cole, M. (1999). Culture-free versus culture-based measures of cognition. . In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 654-664). Cambridge: MIT Press.
- Cravioto, J., DeLicardie, E., & Birch, H. (1966). Nutrition, growth, and neuro-integrative development: an experimental and ecologic study. *Pediatr, 38*, 319-372.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.
- Cueto, S., Leon, J., Guerrero, G., & Munoz, I. (January 2009). *Psychometric characteristics of cognitive development and achievement instruments in Round 2 of Young Lives. Young Lives Technical Note #15.* <<http://www.younglives.org.uk/publications/technical-notes>>.
- Darrah, J., Redfern, L., Maguire, T. O., Beaulne, A. P., & Watt, J. (1998). Intra-individual stability of rate of gross motor development in full-term infants. *Early Human Development, 52*, 169-179.
- Davis, H. L., & Pratt, C. (1996). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47*, 25-31.
- Denham, S., Blair, K., DeMulder, E., Levitas, J., Sawyer, K., Auerbach-Major, S., et al. (2003). Preschool emotional competence: Pathway to social competence. *Child Development, 74*, 238-256.
- Doig, K. B., Macias, M. M., Saylor, C. F., Craver, J. R., & Ingram, P. E. (1999). The Child Development Inventory: A developmental outcome measure for follow-up of the high-risk infant. *Journal of Pediatrics, 135*, 358-362.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428-1446.
- Elliott, C. D. (1996). *The British Ability Scales II*. Windsor, Berkshire: NFER-NELSON Publishing Company.
- Engle, P. L., Black, M. M., Behrman, J. R., Cabral de Mello, M., Gertler, P. J., Kapiriri, L., et al. (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *The Lancet, 369*(9557), 229-242.

- Engle, P. L., Black, M. M., Behrman, J. R., Cabral de Mello, M., Gertler, P. J., Kapiriri, L., et al. (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *The Lancet*, 369(9557), 229-242.
- Evans, G. W. (2003). A multimethodological analysis of cumulative risk and allostatic load among rural children. *Developmental Psychology*, 39(5), 924-933.
- Evans, G. W., & English, K. (2002). The environment of poverty: multiple stressor exposure, psychophysiological stress, and socioemotional adjustment. *Child Development*, 73(4), 1238-1248.
- Ezeilo, B. (1978). Validating Panga Munthu Test and Porteus Maze Test in Zambia. *International Journal of Psychology*, 13, 333-342.
- Feinstein, L. (2003). Inequality in the Early Cognitive Development of British Children in the 1970 Cohort. *Economica*, 70, 73-97.
- Fernald, L. C., & Grantham-McGregor, S. M. (1998). Stress response in children who have been growth retarded since early childhood. *Am J Clin Nutr*, 68, 691-698.
- Fernald, L. C., Neufeld, L. M., Barton, L. R., Schnaas, L., Rivera, J., & Gertler, P. J. (2006). Parallel deficits in linear growth and mental development in low-income Mexican infants in the second year of life. *Public Health Nutr.*, 9(2), 178-186.
- Florenco, C. A. (1988). *Nutrition, health and other determinants of academic achievement and school related behavior of grade one to grade six pupils*. Quezon City: University of the Philippines.
- Frankenburg, W. K. (1985). The Denver approach to early case finding. In W. K. Frankenburg, R. N. Emde & J. W. Sullivan (Eds.), *Early identification of children at risk*. (pp. 135-156). New York: NY: Plenum Press.
- Frankenburg, W. K., Dodds, J., Archer, P., Shapiro, H., & Bresnick, B. (1992). The Denver II: A major revision and restandardization of the Denver Developmental Screening Test. *Pediatrics*, 89(1), 91-97.
- Freeman, H. E., Klein, R. E., Townsend, J. W., & Lechtig, A. (1980). Nutrition and cognitive development among rural Guatemalan children. *Am J Public Health*, 70(12), 1277-1285.
- Fyro, K., & Bodegard, G. (1987). Four-year follow-up of psychological reactions to false positive screening tests for congenital hypothyroidism. *Acta Paediatrica Scand*, 76, 107-114.

- Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3.5-7 years old on a Stroop-like day-night test. *Cognition*, 53, 129-153.
- Gessell, A. (1946). The ontogenesis of infant behavior. In L. Carmichael (Ed.), *Manual of Child Psychology* (pp. 295-331). New York, NY: Wiley.
- Gioia, G. A., Isquith, P. K., Retzlaff, P. D., & Espy, K. A. (2002). Confirmatory factor analysis of the Behavior Rating Inventory of Executive Function (BRIEF) in a clinical sample. *Child Neuropsychology*, 8(4), 249-257.
- Gladstone, M. J., Lancaster, G. A., Jones, A. P., Maleta, K., Mtitimila, E., Ashorn, P., et al. (2008). Can Western developmental screening tools be modified for use in a rural Malawian setting? *Arch Dis Child*, 93(1), 23-29.
- Glascoe, F. P. (2001). Are overreferrals on developmental screening tests really a problem? *Arch Pediatr Adolesc Med*, 155, 54-59.
- Glascoe, F. P. (2005). Screening for developmental and behavioral problems. *Mental Retardation and Developmental Disabilities Research Reviews*, 11(3), 173-179.
- Gottlieb, G. (1991). Experiential canalization of behavioral development: theory. *Developmental Psychology*, 27(1), 4-13.
- Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., Strupp, B., et al. (2007). Developmental potential in the first 5 years for children in developing countries. *Lancet*, 369(9555), 60-70.
- Grantham-McGregor, S. M., Walker, S. P., Chang, S. M., & Powell, C. A. (1997). Effects of early childhood supplementation with and without stimulation on later development in stunted Jamaican children. *Am J Clin Nutr*, 66(2), 247-253.
- Greenfield, P. M. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist*, 52(10), 1115-1124.
- Griffiths, R. (1984). *The abilities of young children*. Amersham: ARICD.
- Grigorenko, E. L., & Sternberg, R. J. (1999). *Assessing cognitive development in early childhood*. Washington D.C.: World Bank.
- Guhn, M., Gadermann, A., & Zumbo, B. D. (2007). Does the EDI measure school readiness in the same way across different groups of children? *Early Education & Development*, 18(3), 453 - 472.

- Guo, G., & Harris, K. M. (2000). The mechanisms mediating the effects of poverty on children's intellectual development. *Demography*, 37(431-47).
- Guthrie, J. F., & Morton, J. F. (1999). Diet-related knowledge, attitudes, and practices of low-income households with children. *Journal of Early Education and Family Review*, 6(3), 26-33.
- Hamadani, J. D., Tofail, F., Yesmin, S., Huda, S. N., Engle, P., & Grantham-McGregor, S. M. (Submitted). Validating family care indicators in Bangladesh. .
- Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, 45, 153-171.
- Handal, A. J., Lozoff, B., Breilh, J., & Harlow, S. D. (2007). Effect of community of residence on neurobehavioral development in infants and young children in a flower-growing region of Ecuador. *Environmental Health Perspectives*, 115, 128-133.
- Harkness, S., & Super, C. M. (1977). Why African children are so hard to test. In L. L. Adler (Ed.), *Issues in Cross-Cultural Research*. New York: The New York Academy of Sciences.
- Harris, D. B. (1963). *Children's drawings as measures of intellectual maturity*. . New York: Harcourt, Brace & World, Inc.
- Hart, B., & Risley, T. R. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology*, 28, 1096-1105.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul Brookes.
- Haussler, I. M., & Marchant, T. (1980). *TEPSI test de desarrollo psicomotor 2-5 anos (TEPSI test of psychomotor development 2-5 years)* (8 ed.). Santiago, Chile: Ediciones Universidad Catolica.
- Heckman, J., & Masterov, D. V. (2005). *The productivity argument for investing in young children*. Chicago: University of Chicago.
- Heo, K. H., Squires, J., & Yovanoff, P. (2008). Cross-cultural adaptation of a pre-school screening instrument: comparison of Korean and US populations. *Journal of Intellectual Disability Research*, 52(3), 195-206.

- Hill, V. (2005). Through the Past Darkly: A Review of the British Ability Scales Second Edition. *Child and Adolescent Mental Health, 10*(2), 87-98.
- Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary via maternal speech. *Child Development, 74*(5), 1368-1378.
- Holding, P. A., Taylor, H. G., Kazungu, S. D., Mkala, T., Gona, J., Mwamuye, B., et al. (2004). Assessing cognitive outcomes in a rural African population: Development of a neuropsychological battery in Kilifi District, Kenya. *Journal of the International Neuropsychological Society, 10*(02), 246-260.
- Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of Hot and Cool Executive Function in Young Children: Age-Related Changes and Individual Differences. *Developmental Neuropsychology, 28*(2), 617 - 644.
- Hooper, S. R., Burchinal, M., Roberts, J. E., Zeisel, S., & Neebe, E. C. (1998). Social and family risk factors for infant development at one year: An application of the cumulative risk model. *Journal of Applied Developmental Psychology, 19*(1), 85-96.
- Howard, D. P., & De Salazar, M. N. (1984). Language and cultural differences in the administration of the Denver Developmental Screening Test. *Child Study Journal, 14*, 1-9.
- Hsieh, C.-C., & Pugh, M. D. (1999). Poverty, income inequality, and violent crime: A meta-analysis of recent aggregate data studies. In I. Kawachi, B. P. Kennedy & R. G. Wilkinson (Eds.), *Income Inequality and Health* (Vol. 1, pp. 278-296). New York: The New Press.
- Jacobs, D. E., Clickner, R. P., Zhou, J. Y., Viet, S. M., Marker, D. A., Rogers, J. W., et al. (2002). The prevalence of lead-based paint hazards in U.S. housing. *Environmental Health Perspectives, 110*(10), A599-606.
- Jamison, D. (1986). Child malnutrition and school performance in China. *Journal of Development Economics, 20*, 299-309.
- Janus, M., & Offord, D. (2007). Development and psychometric properties of the Early Developmental Inventory (EDI): a measure of children's school readiness. *Canadian Journal of Behavioral Science, 39*(1), 1-22.

- Johnson, M. H. (1998). The neural basis of cognitive development. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of Child Psychology, 5th edition. Volume 2: Cognition, Perception and Language*. (pp. 1-50). New York: John Wiley.
- Johnson, S., & Marlow, N. (2006). Developmental screen or developmental testing? *Early Human Development*, 82, 173-183.
- Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology Review*, 17(3), 213-233.
- Kagan, S. L., & Britto, P. (2005). *Report on "Going Global"*. . New York: UNICEF.
- Kagan, S. L., Britto, P. R., & Engle, P. L. (2005). Early learning standards: What can America learn? What can America teach? . *Phi Delta Kappan*, 87(3), 205-208.
- Kagitcibasi, C., Sunar, D., & Bekman, S. (2001). Long-term effects of early intervention: Turkish low-income mothers and children. *Journal of Applied Developmental Psychology*, 22, 333-361.
- Kariger, P. K., Stoltzfus, R. J., Olney, D., Sazawal, S., Black, R., Tielsch, J. M., et al. (2005). Iron deficiency and physical growth predict attainment of walking but not crawling in poorly nourished Zanzibari infants. *Journal of Nutrition*, 135(4), 814-819.
- Karns, J. T. (2001). Health, nutrition and safety. In G. Bremner & A. Fogel (Eds.), *Handbook of Infant Development* (pp. 693-725). Malden, MA: Blackwell.
- Kathuria, R., & Serpell, R. (1998). Standardization of the Panga Munthu Test—a nonverbal cognitive test developed in Zambia. *Journal of Negro Education*, 67, 228–241.
- Koch, R., Lewis, M. T., & Quinones, W. (1998). Homeless: Mothering at rock bottom. In C. G. Coll & J. L. Surrey (Eds.), *Mothering against the odds: Diverse voices of contemporary mothers* (pp. 61-84). New York, NY, USA: The Guilford Press.
- Kovas, Y., Haworth, C. M. A., Dale, P. S., & Plomin, R. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development*, 288, 1-144.
- Kuhn, D., & Siegler, R. S. (1998). *Handbook of Child Psychology, Fifth edition. Volume 2: Cognition, Perception and Language*.
- Kuklina, E. V., Ramakrishnan, U., Stein, A. D., Barnhart, H. H., & Martorell, R. (2004). Growth and diet quality are associated with the attainment of walking in rural Guatemalan infants. *Journal of Nutrition*, 134(12), 3296-3300.

- Lansdown, R. G., Goldstein, H., Shah, P. M., Orley, J. H., Di, G., Kaul, K. K., et al. (1995). Culturally appropriate measures for monitoring child development at family and community level: A WHO collaborative study. *Bulletin of the World Health Organization*, 74(3), 283-290.
- Lasky, R. E., Klein, R. E., Yarbrough, C., Engle, P. L., Lechtig, A., & Martorell, R. (1981). The relationship between physical growth and infant behavioral development in rural Guatemala. *Child Dev*, 52(1), 219-226.
- Leerkes, E., & Crockenberg, S. (2003). The Impact of Maternal Characteristics and Sensitivity on the Concordance Between Maternal Reports and Laboratory Observations of Infant Negative Emotionality. *Infancy*, 4(4), 517-539.
- Lerner, R. M. (1998). Theories of human development: contemporary perspectives. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology, 5th edition. Volume 1: Theoretical models of human development*. New York: John Wiley & Sons, Inc.
- Love, J. M., Kisker, E. E., Ross, C., Constantine, J., Boller, K., Tarullo, L. B., et al. (2005). The effectiveness of Early Head Start for children and their parents: Lessons for policy and programs. *Developmental Psychology*, 41(6), 885-901.
- Malda, M., Van De Vijver, F. J. R., Srinivasan, C., Transler, C., Sukumur, P., & Rao, K. (2008). Adapting a cognitive test for a different culture: An illustration of qualitative procedures. *Psychology Science Quarterly*, 50(4), 451-468.
- Marchman, V. A., & Martine-Sussmann, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both English and Spanish. *J Speech Lang Hear Res*, Oct;45(5), 983-997.
- McCall, R. B. (1981). Nature-nurture and the two realms of development: A proposed integration with respect to mental development. *Child Development*, 52, 1-12.
- McCartney, K., Dearing, E., Taylor, B. A., & Bub, K. L. (2007). Quality child care support the achievement of low-income children: Direct and indirect pathways through caregiving and the home environment. *Journal of Applied Developmental Psychology*, 28(5-6), 411-426.
- Miesels, S. J., & Atkins-Burnett, S. (2000). The elements of early childhood assessment. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed.). Cambridge: Cambridge University Press.

- Monckeberg, F. (1972). Malnutrition and mental capacity. In P. A. H. Organization (Ed.), *Nutrition, the nervous system and behaviour, Scientific Publication No 251* (pp. 48-54). Washington D.C.: PAHO.
- Montie, J. E., Xiang, Z., & Schweinhart, L. J. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. *Early Childhood Research Quarterly, 21*, 313-331.
- Moock, P. R., & Leslie, J. (1986). Childhood malnutrition and schooling in the Teri region of Nepal. *Journal of Development Economics, 20*, 33-52.
- Mpofu, E. (1995). Antecedents of children's performance on class inclusion tasks: Some Zimbabwean evidence. *International Journal of Psychology, 30*(1), 19-33.
- Mulatu, M. S. (1995). Prevalence and risk factors of psychopathology in Ethiopian children. *J Am Acad Child Adolesce Psychiatry, 34*(1), 100-109.
- Mustard, J. F., & Young, M. E. (2007). Measuring child development to leverage ECD policy and investment. In M. E. Young & L. M. Richardson (Eds.), *Young early child development: From measurement to action*. Washington, DC: The World Bank.
- Mwamwenda, T. S., & Mwamwenda, B. B. (1989). Assessing Africans' cognitive development: Judgement or judgement and explanation. *Journal of Genetic Psychology, 151*(2), 245-254.
- Mwaura, P. A. M., Sylva, K., & Malmberg, L. (2008). Evaluating the Madrasa preschool programme in East Africa: A quasi-experimental study. *International Journal of Early Years Education, 16*(3), 237-255.
- Neisser, U., Boodoo, G., Bouchard Jr., T. J., Boykin, A. W., Brody, N., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*(2), 77-101.
- Nelson, C. A. (2000). The neurobiological basis of early intervention. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed.). Cambridge: Cambridge University Press.
- Nerlove, M. (1974). Household and economy: Toward a new theory of population and economic growth. *The Journal of Political Economy, 82*(2), S200-S218.
- NICHD Early Child Care Research Network. (2003). Do children's attention processes mediate the link between family predictors and school readiness? *Developmental Psychology, 39*, 581-593.

- Noble, K. G., Tottenham, N., & Casey, B. J. (2005). Neuroscience Perspectives on Disparities in School Readiness and Cognitive Achievement. *The Future of Children, 15*(1), 71-89.
- Ogunnaike, O. A., & Houser Jr., R. F. (2002). Yoruba toddlers' engagement in errands and cognitive performance on the Yoruba Mental Subscale. *International Journal of Behavioral Development, 26*(2), 145-153.
- Oiberman, A. (2005). Evaluación de la inteligencia en bebés argentinos: Escala Argentina de Inteligencia Sensorio-motriz. *Revista Argentina de Clínica Psicológica, 14*(3), 213-218.
- Oiberman, A. (2006). Resiliencia y factores de protección en bebés vulnerables. Aplicación de la Escala Argentina de Inteligencia Sensoriomotriz. . *Acta Psiquiátrica y Psicológica de America Latina, 52*(1), 19-25.
- Paine, P., Dorea, J. G., Pasquali, L., & Monteiro, A. M. (1992). Growth and cognition in Brazilian schoolchildren: a spontaneously occurring intervention study. *International Journal of Behavioral Development, 15*(2), 169-183.
- Payne, K. T., & Taylor, O. L. (2002). Multicultural influences on human communication. In G. H. Shames & N. B. Anderson (Eds.), *Human communication disorders. An introduction* (6 ed.). Boston, MA:: Ally & Bacon.
- Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development, 78*(4), 1255-1264.
- Pollitt, E. (2001). The developmental and probabilistic nature of the functional consequences of iron-deficiency anemia in children. *J. Nutr., 131*(2), 669S-675S.
- Pollitt, E., & Triana, N. (1999). Stability, predictive validity, and sensitivity of mental and motor development scales and pre-school cognitive tests among low-income children in developing countries. *Food and Nutrition Bulletin, 20*(1), 45-52.
- Powell, C., & Grantham-McGregor, S. (1980). The associations between nutritional status, school achievement and school attendance in 12-year old children at a Jamaican school. *West Indian Medical Journal, 29*, 247-253.
- Powell, C. A., & Grantham McGregor, S. (1985). The ecology of nutritional status and development in young children in Kingston, Jamaica. *Am J Clin Nutr, 41*(6), 1322-1331.
- Reynell, J. (1990). *Reynell Developmental Language Scales*. Los Angeles: Western Psychological Services.

- Rimm-Kaufman, S. E., Pianta, R. B., & Cox, M. J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly, 15*, 147-166.
- Rodriguez, S. (1996). *Escala de evaluacion del desarrollo psicomotor: 0 a 24 meses*. (12 ed.). Santiago, Chile: Galdoc.
- Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and Cognition, 52*, 326-333.
- Rubin, K., Hemphill, S., Chen, X., Hastings, P., Sanson, A., & Coco, A. (2006). A cross-cultural study of behavioral inhibition in toddlers: East-West-North-South. *International Journal of Behavioral Development, 30*(3), 219-226.
- Rubin, K., Hemphill, S., Chen, X., Hastings, P., Sanson, A., & LoCoco, A. (2006). Parenting Beliefs and Behaviors: Initial Findings From the International Consortium for the Study of Social and Emotional Development (ICSSSED). *Parenting beliefs, behaviors, and parent-child relations: A cross-cultural perspective* (pp. 81-103). New York: Psychology Press.
- Rutter, M. (1979). Protective factors in children's responses to stress and disadvantage. In M. W. Kent & J. E. Rolf (Eds.), *Primary prevention of psychopathology* (Vol. 3, pp. 49-74). Hanover, NH: University Press of New England.
- Rydz, D., Shevell, M., Majnemer, A., & Oskoui, M. (2005). Topical review: Developmental screening. *Journal of Child Neurology, 20*(4), 4-21.
- Saarni, C., Mumme, D. L., & Campos, J. J. (1998). Emotional development: action, communication, and understanding. In N. Eisenberg (Ed.), *Handbook of Child Psychology, 5th edition. Volume 3: Social, Emotional, and Personality Development*. (pp. 237-310). New York: John Wiley & Sons, Inc.
- Sameroff, A. J., Seifer, R., Baldwin, A., & Baldwin, C. (1993). Stability of intelligence from preschool to adolescence: The influence of social and family risk factors. *Child Development, 64*(1), 80-97.
- Sameroff, A. J., Seifer, R., Barocas, R., Zax, M., & Greenspan, S. (1987). Intelligence quotient scores of 4-year-old children: social- environmental risk factors. *Pediat, 79*(3), 343-350.
- Scarborough, A. A., Hebbeler, K. M., Simeonsson, R. J., & Spiker, D. (2007). Caregiver descriptions of the developmental skills of infants and toddlers entering early intervention services. *Journal of Early Intervention, 29*(3), 207-227.

- Schady, N. (2006). Early childhood development in Latin American and the Caribbean. *Economía*, 6.2, 185-225.
- Schneider, W., & Bjorklund, D. F. (1998). Memory. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of Child Psychology, 5th edition. Volume 2: Cognition, Perception and Language*. (pp. 1-50). New York: John Wiley.
- Sebate, M. (2000). *Report on the standardisation of the Grover-Counter Scale of Cognitive Development*. Pretoria, South Africa: Human Services Research Council.
- Sen, A. (1999). *Development as Freedom*: Knopf.
- Serpell, R., & Jere-Folotiya, J. (2008). Developmental assessment, cultural context, gender and schooling in Zambia. *International Journal of Psychology*, 43(2), 1-9.
- Shonkoff, J. P., & Marshall, P. C. (2000). Neurological basis of developmental vulnerability. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed.). Cambridge: Cambridge University Press.
- Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From Neurons to Neighborhoods: The Science of Early Childhood. Development Committee on Integrating the Science of Early Childhood Development*. Washington D.C.: National Academy Press.
- Sigman, M., Neumann, C., Jansen, A. A., & Bwibo, N. (1989). Cognitive abilities of Kenyan children in relation to nutrition, family characteristics, and education. *Child Dev*, 60(6), 1463-1474.
- Silva, P. (1986). A comparison of the predictive validity of the Reynell Developmental Language Scales, the Peabody Picture Vocabulary Test and the Stanford–Binet Intelligence Scale. *British Journal of Educational Psychology* 56, 201-204.
- Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early Childhood Assessment: Why, What, and How*. Washington, D.C.: The National Academies Press.
- Solarsh, B., & Alant, E. (2006). The challenge of cross-cultural assessment—The Test of Ability To Explain for Zulu-speaking Children. *Journal of Communication Disorders*, 39, 109-138.
- Squires, J. K., Potter, L., Bricker, D. D., & Lamorey, S. (1998). Parent-completed developmental questionnaires: Effectiveness with low and middle income parents. *Early Childhood Research Quarterly*, 13(2), 345-354.

- Stoltzfus, R., Kvalsvig, J., Chwaya, H., Montresor, A., Albonico, M., Tielsch, J., et al. (2001). Effects of iron supplementation and anthelmintic treatment on motor and language development of preschool children in Zanzibar: double blind, placebo controlled study. *BMJ*, *323*, 1-8.
- Tamis-LeMonda, C.S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Dev.*, *72*, 748-767.
- Thelen, E. (2000). Grounded in the world: Developmental origins of the embodied mind. *Infancy*, *1*(1), 3-28.
- Thompson, R. A., & Raikes, H. A. (2006). The social and emotional foundations of school readiness. In J. Knitzer, R. Kaufmann & D. Perry (Eds.), *Early childhood mental health*. Baltimore, MD: Paul H. Brookes Publishing Co.
- Tluczek, A., Mischler, E., Farrell, P. M., & et.al. (1992). Parents' knowledge of neonatal screening and response to false-positive cystic fibrosis testing. *J Dev Behav Pediatr.*, *13*, 181-186.
- Van De Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: some practical guidelines. *European Psychologist*, *1*(2), 89-99.
- van Widenfelt, B. M., Treffers, P. D. A., de Beurs, E., Siebelink, B. M., & Koudijs, E. (2005). Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clinical Child and Family Psychology Review*, *8*(2), 135-147.
- Vazir, S., & Kashinath, K. (1999). Influence of the ICDS on psychosocial development of rural children in Southern India. *Journal of the Indian Academy of Applied Psychology*, *25*(1), 11-24.
- Wachs, T. (1993). Family environmental influences and development: Illustrations from the study of undernourished children.(pp. 245-268). *Families, risk, and competence*. Mahway, NJ: Lawrence Erlbaum Associates Publishers.
- Wachs, T., & Desai, S. (1993). Parent-report measures of toddler temperament and attachment: Their relation to each other and to the social microenvironment. *Infant Behavior and Development*, *16*(3), 391-396.

- Wachs, T., Sigman, M., Bishry, Z., & Moussa, W. (1992). Caregiver child interaction patterns in two cultures in relation to nutritional intake. *International Journal of Behavioral Development, 15*(1), 1-18.
- Wachs, T. D., Sigman, M., Bishry, Z., Moussa, W., Jerome, N., Neumann, C., et al. (1992). Caregiver child interaction patterns in two cultures in relation to nutritional intake. *International Journal of Behavioral Development, 15*, 1-18.
- Wagstaff, A., Bustreo, G., Bryce, J., Claeson, M., & WHO (2004). Child Health: Reaching the Poor. *American Journal of Public Health, 94*(5), 726-736.
- Walker, S. P., Wachs, T. D., Gardner, J. M., Lozoff, B., Wasserman, G. A., Pollitt, E., et al. (2007). Child development: risk factors for adverse outcomes in developing countries. *Lancet, 369*(9556), 145-157.
- Wamboldt, F. S., Ho, J., Milgrom, H., Wamboldt, M. Z., Sanders, B., Szeffler, S. J., et al. (2002). Prevalence and correlates of household exposures to tobacco smoke and pets in children with asthma. *Journal of Pediatrics, 141*(1), 109-115.
- Welsh, M. C., Friedman, S. L., & Spieker, S. J. (2006). Executive functions in developing children: Current conceptualizations and questions for the future. In K. McCartney & D. Phillips (Eds.), *Blackwell handbook of early childhood development*. London: Blackwell.
- Werner, E. E. (2000). Protective factors and individual resilience. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed.). Cambridge: Cambridge University Press.
- WHO, Multicentre Reference Study Group. (2006a). Assessment of sex differences and heterogeneity in motor milestone attainment among populations in the WHO Multicentre Growth Reference Study. *Acta Paediatrica, S450*, 55-75.
- WHO, Multicentre Reference Study Group (2006b). WHO motor development study: Windows of achievement for six gross motor development milestones. *Acta Paediatrica, S450*, 86-95.
- Woodward, A. L., & Markman, E. M. (1998). Early word learning. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of Child Psychology, 5th edition. Volume 2: Cognition, Perception and Language*. (pp. 1-50). New York: John Wiley.

APPENDIX A: Details of published and normed measures

Summary of options for published and normed measures

	Includes cognition and other domains	Cognition only	Language only	Socio-emotional only
Direct assessment	-Bayley Scales (A) -British Ability Scales (A) -Denver (S) -Griffiths (A) -McCarthy (A) -Stanford Binet (A)	-Kaufman ABC (A) -Leiter (A) -WPPSI (A) -WISC (A) -Woodcock-Johnson (A)	-Peabody Picture Vocabulary Test (PPVT or TVIP) (A) -Preschool Language Scale (A) -Reynell Language Development (A)	
Ratings & reports	-Ages and Stages (S)		-MacArthur CDI (A)	-Achenbach CBCL (A) -Infant and Toddler Socio-Emotional Assessment (A) -Strengths and Difficulties Questionnaire (S)

** (S) indicates that the assessment is a screening tool; (A) indicates that it is an assessment of abilities.

Note: observational methods are not reviewed in this appendix. Also, Pegboard is included in the Appendix below, but not in the Chapter 8 above.

Achenbach Child Behavior Checklist (CBCL) 1 1/2-5 and Caregiver-Teacher Report Form (C-TRF)

CBCL:
-Domain: Socio-emotional
-Type of assessment: Abilities
-Mode of assessment: Ratings/Reports

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>Assess behavioral and emotional problems in children.</p> <p>Age range: 1 year, 6 months to 5 years, 11 months</p>	<p>The normative sample was derived from a national probability sample (the National Survey) collected in 1999 by the Institute for Survey Research.</p> <p>No mention of international norms.</p>	<p>Parent or caregiver is respondent for CBCL 1 1/2-5; Teacher or caregiver in preschool or daycare setting is respondent for C-TRF. These measures are usually administered as written questionnaires. If respondent has reading difficulties, it can be interviewer-administered, one on one.</p>	<p>Respondents should have at least fifth-grade reading skills. Since these are usually administered as written questionnaires, little specific training is required for administration.</p>	<p>10-15 minutes. Cost is \$250 for Pre-school Computer-scoring Starter Kit (50 CBCL/12-5-LDS forms, 50-TRF forms, ADM with Ages 11/2-5 module for the CBCL-LDS & C-TRF, and manual). Cost for Hand-Scoring Starter Kit is \$150.</p>

Publisher: University of Vermont, Research Center for Children, Youth, and Families
 ASEBA (Achenbach System of Empirically Based Assessment). 1 South Prospect Street
 Burlington, VT 05401-3456 Tel: 802-264-6432; Fax: 802-264-6433
 Website: www.aseba.org

Ages and Stages Questionnaire (ASQ)

ASQ:

-Domain: Comprehensive

-Type of assessment: Screening

-Mode of assessment: Ratings/Reports

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
To screen infants and young children for developmental delays during the first 5 years of life. The assessment covers five key developmental areas: communication, gross motor, fine motor, problem solving, and personal-social. 4 to 60 months	No norms available. Information on research on the test (particularly on validity and reliability) can be found at http://www.brookespublishing.com/store/books/bricker-asq/asq-introduction.pdf .	Parent-Teacher self-report. Parents complete the 30-item questionnaires at designated intervals, assessing children in their natural environments.	Not specified. According to publishers, professionals convert parents' responses of yes, sometimes, and not yet to color-coded scoring sheets, enabling them to quickly determine a child's progress in each developmental area. The ASQ User's Guide then offers guidelines for determining whether children are at high or low risk in the various domains.	10 - 15 minutes. Cost is \$199 for the "complete ASQ system," which includes 19 color-coded, questionnaires, 19, age-appropriate scoring sheets, 1 storage box, and the ASQ User's Guide Questionnaires available in English, Spanish, French, and Korean.

Publisher: Brookes Publishing Co. P.O. Box 10624 Baltimore, MD 21285-0624
<http://www.brookespublishing.com>

Bayley Scales of Infant Development (BSID-I, 1st edition; BSID-II, 2nd edition; BSID-III, 3rd edition)

BSID:
-Domain: Comprehensive
-Type of assessment: Abilities
-Mode of assessment: Direct

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>Assess the developmental status of infants and children in a wide range of domains. The primary value of the test is in diagnosing developmental delay and planning intervention strategies.</p> <p>1-42 months</p>	<p>Norms were developed from a sample of 1700. The sample was stratified on age, gender, race, geographic region and parent education. No mention of international norms.</p>	<p>Administered directly, one-on-one</p> <p>Infant or young child is respondent. Respondent is asked to perform tasks. Smiles when examiner smiles, eyes following ring of motion, removes lid from box after watching examiner put toys inside.</p>	<p>Must be administered by someone trained to administer the test and experience testing young children</p>	<p>25-35 minutes for children under 15 months and 60 minutes for children over 15 months.</p> <p>Cost is \$1,045 for Bayley-III Comprehensive Kit and Screening Test Kit Combo</p>

Publisher: Harcourt Assessment, 19500 Bulverde Rd., San Antonio TX 78259
 Phone: 1-800-872-1726; Website: www.psychcorp.com

British Ability Scales (BAS)

BAS:

-Domain: Comprehensive

-Type of assessment: Abilities

-Mode of assessment: Direct Assessment

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>Measures core (verbal, visual/spatial, and non-verbal) as well as subscales for differential abilities and in the older group, achievement tests . Purpose is the assessment of particular cognitive abilities linked to developing understandings and supporting interventions rather than categorization of children. This facilitates the movement away from the restrictive practice of generating broad and general assessment information across a range of cognitive abilities with a focus on categorisation rather than intervention.</p> <p>2.6 years through 17.2 years (Early Years battery)</p>	<p>Nationally representative sample in UK in 1995 of 1689 children</p>	<p>Series of test items given to child by trained tester. Step and start points defined based on ceiling and basal rules. Can use sub-tests separately.</p>	<p>Examiners should have thorough understanding of the administration and scoring procedure and formal training in assessment.</p>	<p>Depends on number of subtests administered. The Core battery is 4-6 tests, depending on child age. Manual, test kit for the Early Years</p> <p>Costs ~650 British Pounds.</p>

Publisher: GL Assessment. <http://shop.gl-assessment.co.uk/home.php?cat=303>

Denver Developmental Materials II (formerly DDST)

Denver:
-Domain: Comprehensive
-Type of assessment: Screening
-Mode of assessment: Direct Assessment

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>This is a surveillance and monitoring instrument used by professionals or trained paraprofessionals to determine if a child's development is within the normal range. The results are not diagnostic. The DENVER II is designed to reflect the development of a broad range of heterogeneous skills in a minimum amount of time. As such it is not designed to measure any single construct such as intelligence, motor functioning, or social skills.</p> <p>Birth to 6 years</p>	<p>To standardize this test, the sample of over 2,000 children, representing a broad spectrum of children, was representative of the Colorado population (1980 census). This sample has relatively minor demographic differences between it and the U.S.</p>	<p>Administered directly, one-on-one</p> <p>Infant or young child is respondent. Respondent is asked to perform tasks.</p>	<p>According to the publishers, "anyone who works well with children and meticulously follows directions for administration can be a screener." However, training is required (see "time needed and cost").</p>	<p>Between 10 and 20 minutes to administer and interpret the test, depending on child's age and cooperation.</p> <p>Cost is \$90 for materials (100 DENVER II forms, DENVER II Training Manual, and DENVER II kit). Additional set of two training videos is \$410 or \$185 for one-week rental. Training sessions also held 3 X / year in Denver for \$395 per person.</p>

Publisher: Denver Developmental Materials
P.O. Box 371075
Denver, CO 80237-5075
(303) 355-4729
(800) 419-4729
(303) 355-5622 (fax)
Website: <http://www.denverii.com/>

Infant and Toddler Socio-Emotional Assessment (ITSEA, or BITSEA – brief form)

ITSEA:
-Domain: Socio-emotional
-Type of assessment: Abilities
-Mode of assessment: Ratings/Reports

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>Developed to assess socio-emotional problems and competencies.</p> <p>12-48 months</p>	<p>Psychometric work and norms development were conducted with a Community Survey sample of infants and toddlers. 1,280 families participated in the study. The sample was roughly balanced in terms of gender.</p>	<p>One-on-one administration; parents or other caregivers are the respondents.</p>	<p>Little specific training is required. It can be administered as a questionnaire or an interview. If interview-administered, interviews should be trained to administer the test in a standard manner.</p>	<p>It takes approximately 20 to 30 minutes to complete independently as a questionnaire, or 35 to 40 minutes to complete as an interview.</p> <p>Cost is \$150 for ITSEA Parent Forms 10, Child Care Provider Forms 10, ITSEA Manual</p>

Publisher: Harcourt Assessment
 19500 Bulverde Rd. San Antonio TX 78259.
 Phone: 1-800-211-8378
 Website: www.psychcorp.com

Has also been translated into Chinese.

Kaufman Assessment Battery for Children (Kaufman ABC)

K-ABC:
-Domain: Cognition
-Type of assessment: Abilities
-Mode of assessment: Direct

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>For comprehensive assessment of preschool children – included multiple measures of cognition and language</p> <p>2 years 6 months through 12 years 6 months.</p>	<p>Norm-referenced. The norm sample for the KABC-II (2nd edition) closely matches 2001 census data with respect to race/ethnicity, gender, region, SES, special education status.</p>	<p>One-on-one administration</p> <p>Child is the respondent, who responds to requests made by the examiner. The child is required to give a verbal response, point to a picture, build something, etc.</p>	<p>Training is required. Examiner should be well-versed in psychology and individual intellectual assessment, who has studied carefully the KABC materials. "...those who are not permitted to administer existing intelligence scales do not ordinarily possess the skills to be K-ABC examiners." (Kaufman & Kaufman, 1983)</p>	<p>35 minutes (at 2 years, 6 months) to 75-85 minutes (at age 7 and above).</p> <p>Cost is \$724.99 (4 easels, one manual, all stimulus and manipulative materials, 25 record forms, and briefcase.) (from 2003 KABC-II brochure)</p>

Publisher: Pearson Assessments
 Tel: (800) 627-7271
 Website: www.ags.pearsonassessments.com

Leiter-R or Leiter International Performance Scale

Leiter:

-Domain: Cognition

-Type of assessment: Abilities

-Mode of assessment: Direct

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>This is a nonverbal measure of the child's intellectual ability across several domains. Because it is nonverbal, it is suitable for children and adolescents that are cognitively delayed, disadvantaged, nonverbal or non-English speaking, ESL, speech, bearing or motor impaired, ADHD, autistic, and TBI.</p> <p>2 years to 20 years 11 months</p>	<p>Standardized on 1,719 "typical" children and adolescents, and 692 atypical children (representing 9 clinical groups) ages 2-0 to 20-11, using a national stratification plan. Nationally-representative proportions of Caucasian (non-Hispanic), Hispanic-American, African-American, Asian-American and Native American children were included.</p>	<p>One-on-one administration</p> <p>Child is the respondent, who responds to requests made by the examiner. Tests are described as "game-like".</p>	<p>Training is required for the standardization procedures of the task. According to the publishers, "It does not require a spoken or written word from the examiner or the child. The easy game-like administration holds the child's interest and is easily administered." Training videos and manuals are available.</p>	<p>25-40 minutes. \$895.00 (Manual; 3 Easel Books; VR Response Cards; AM Response Cards; Manipulatives; one package each of the VR and AM Record Forms (20 per package); the Attention-Sustained Booklets; the Rating Scale Booklets - Parent, Teacher and Self (the Examiner Rating Scale is included in the Record Booklet), the Growth Profile Booklet, and rolling backpack)</p>

Publisher: Stoelting Co.
 620 Wheat Lane
 Wood Dale, IL 60191 USA
 Phone: (630) 860-9700
 FAX: (630) 860-9775
 Website: <http://www.stoeltingco.com/>

MacArthur Child Development Inventory (CDI)

CDI:

-*Domain: Language*

-*Type of assessment: Abilities*

-*Mode of assessment: Ratings/Reports*

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>(1) The CDI/Words and Gestures (for 8 - 16 month-olds) assesses vocabulary comprehension, vocabulary production, and the use of gestures. (2) The CDI/Words and Sentences (for 16-30 month olds) assesses vocabulary production and number of aspects of grammatical development, including sentence complexity and mean length of child's longest utterances.</p> <p>8 months through 2 years, 6 months</p>	<p>Sample included 671 families and 1,142 toddlers, with approximately equal numbers of boys and girls in each age range. Race/ ethnicity: 86.9% white; 4% black; 2.9% Asian/pacific islander; 6.2% other. Education: 53.3% had college degree; 24.3% had some college; 17.9% had a high school diploma; and 4.5% had some high school or less. No mention of international norms.</p>	<p>The CDI is completed by the child's parent. For the CDI/Words and Gestures, for the "Words" portion, the parents are asked questions about whether the child is responding to language, or comprehends or uses particular words from a provided list. For the CDI Words and Sentences, parents are asked to fill in a vocabulary production checklist, followed by other questions about use of language.</p>	<p>Some training is required of an examiner.</p>	<p>20 to 40 minutes.</p> <p>Cost is \$90.00 for complete set of forms (package of 20) and user's guide. Technical manual is \$59.95.</p>

Publisher: Brookes Publishing Co.

P.O. Box 10624

Baltimore, MD 21285-0624

Website: <http://www.brookespublishing.com/>

McCarthy Scales of Children's Abilities (MSCA)

MSCA:

-Domain: Comprehensive

-Type of assessment: Abilities

-Mode of assessment: Direct

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>Used to assess cognitive and motor development in children. Useful as an aid in screening and diagnostic decisions. According to SERS (http://www.ctserc.org/aboutus/), The scales "yield a general cognitive index, as well as five subscale scores. The tests are useful in their specification of strengths and weaknesses, and they are appealing to children, but the validity of some scores has been questioned. The norms are also out dated." (It was published in 1972.)</p> <p>2.6 to 8.6 years.</p>	<p>The test was standardized on a sample of 1,032 children stratified by race, geographic region, father's occupational status, and urban-rural residency (in accordance with 1970 U.S. Census data). "Exceptional" children were excluded from the standardization sample.</p>	<p>One-on-one administration directly with child.</p>	<p>Training of assessors is required for the standardized administrative procedures of the task.</p>	<p>45 minutes for children under 5; 1 hour for older children.</p> <p>Cost is \$599 for stimulus and manipulable materials, Manual, 25 Record Forms, 25 Drawing Booklets, and Attache Case.</p>

Publisher: Brookes Publishing Co.
P.O. Box 10624
Baltimore, MD 21285-0624
Website: <http://www.brookespublishing.com/>

Pegboard

Pegboard:

-Domain: Fine motor

-Type of assessment: Abilities

-Mode of assessment: Direct

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
Tests fine motor skills, which are a reliable indicator of children's cognitive skills	Normed for children beginning at 6 years but not for younger children	This test is used extensively to evaluate lateralized brain damage in adults, adolescents, and children whenever manual dexterity is at issue. Consisting of 25 holes with randomly positioned slots, this test requires more complex visual-motor coordination than most pegboards. Pegs with a key on one side must be rotated to match the hole before they can be inserted.	Requires some training of administrator.	5 minutes

Publisher: Psychological Assessment Resources or Stoetling (for Purdue Pegboard). May also be manufactured in country.

Preschool Language Scale (PLS-4)

PLS-4:

-Domain: Language

-Type of assessment: Abilities

-Mode of assessment: Direct

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>Used to identify children who have a language disorder or delay. Provides two core language subscales and three supplemental assessments: The Language Sample Checklist; the Articulation Screener; the Caregiver Questionnaire.</p> <p>Birth through 6 years, 11 months</p>	<p>Standardized based on demographic information obtained from the 2000 census. The sample, which was stratified by parent education, geographic region, and race, included 1,564 children between the ages of 2 days and 6 years, 11 months. The distribution of boys and girls was roughly equal. The sample was roughly comparable to 2000 Census data in terms of region, race/ethnicity, and education levels of primary caregivers.</p>	<p>The child is the respondent. Tasks vary depending on construct and difficulty level. For auditory comprehension, early tasks include whether a child glances at the person speaking to her. Expressive communication tasks include young infant' ability to suck/swallow or vary the pitch , timbre, or length of a cry.</p>	<p>Training is needed; the assessment is usually administered by speech-language pathologists, early childhood specialists, psychologists, educational diagnosticians, properly trained paraprofessionals, and others who have experience and training in assessment.</p>	<p>Birth to 11 months: 20-40 minutes; 12 months to 3 years, 11 months: 30 - 40 minutes; 4 years to 6 years, 11 months: 35 - 45 minutes.</p> <p>Cost is \$275 for kit (Examiner's Manual, Picture Manual, 15 Record Forms, and 23 Manipulatives)</p>

Publisher: Harcourt Assessment
 19500 Bulverde Rd. San Antonio TX 78259.
 Phone: 1-800-872-1726
 Website: www.psychcorp.com

Peabody Picture Vocabulary Test (PPVT)/Test de Imagenes Peabody (TVIP)

PPVT/TVIP:
-Domain: Language
-Type of assessment: Abilities
-Mode of assessment: Direct

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>A test of listening comprehension for the spoken word in standard English. Has two purposes 1) a measure of receptive (hearing) vocabulary; 2) screening test for verbal ability, or as an element in a comprehensive battery of cognitive processes.</p> <p>2 years 6 months through 90+ years</p>	<p>Has reference norms in English and Spanish. Major limitation to Spanish norms is that the sample was small and homogenously high SES</p>	<p>Child or adult is respondent. Picture plates are presented. Each picture plate presents 4 numbered cards simultaneously, and only one card represents a simultaneous word pictorially. Respondent is asked to identify verbally or behaviorally which card represents the stimulus word.</p>	<p>Can be administered by someone familiar with testing and scoring materials (formal training in psychometrics not required). Potential difficulty in scoring test because requires in field calculations.</p>	<p>11-15 minutes.</p> <p>Cost is \$379.99 for complete kit (forms A & B)</p>

Publisher: Pearson Assessments
 Tel: (800) 627-7271
 Website: <http://ags.pearsonassessments.com>

Reynell Language Development Scales

Reynell:

-Domain: Language

-Type of assessment: Abilities

-Mode of assessment: Direct

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>The Verbal Comprehension Scale measures receptive language skills. Two parallel but separately normed versions are provided--one for children who can respond orally, the other for children who can respond only by pointing.</p> <p>The Expressive Language Scale assesses expressive language skills, using three sets of items: Structure, Vocabulary, and Content.</p> <p>1-6 years.</p>	<p>Norms, based on a sample of more than 600 children, reflect U.S. demographics in terms of geographic region, ethnic composition, and parental education.</p> <p>The test provides standard scores, percentiles, and developmental age scores.</p>	<p>The child is the respondent. 134 items are administered directly to the child, using pictures, toys and puppets. Can be done at home; usually is conducted in clinical or school setting.</p>	<p>Training is necessary. Assessors should have college level education and experience with young children.</p>	<p>30 minutes.</p> <p>Cost is \$549 for Complete Test Kit (Includes a complete Set of Stimulus Materials; 10 Test Booklets; 1 Manual, all in a sturdy carrying case)</p>

Western Psychological Services

12031 Wilshire Blvd. Los Angeles, CA 90025-1251

Telephone: (800) 648-8857 - FAX: (310) 478-7838

http://portal.wpspublish.com/portal/page?_pageid=53,69748&_dad=portal&_schema=PORTAL

Stanford Binet Intelligence Scale

Stanford Binet:
-Domain: Cognition
-Type of assessment: Abilities
-Mode of assessment: Direct

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>This test is used to study the development of cognitive skills of individuals. The measure contains 15 subtests that assess mental abilities in four areas: Verbal reasoning, Abstract Visual Reasoning, Quantitative Comprehension and Short-term Memory.</p> <p>Age 2 upwards</p>	<p>The sampling design for the standardization sample, which was used for all of the subtests, was based on five variables, corresponding to 1980 Census data. The variables were geographic region, community size, ethnic group, age, and gender.</p>	<p>The child is the respondent. The SB-IV utilizes basals and ceilings within each subtest, based on sets of four items.. A child is never administered all of the subtests. Guidelines for the tests to be administered are based on the entry level of the examinee.</p>	<p>The administrator should be familiar with the instrument and sensitive to the needs of the examinee. The tester should follow standard procedures, establish adequate rapport between the examiner and the examinee, and correctly score the examinee's responses.</p>	<p>Time limits are not used.</p> <p>Cost is \$937 for Complete Test Kit (Includes 3 Item Books, Examiner's Manual, Technical Manual, Child Card, Layout Card, Manipulatives Kit and Storage Box, and 25 Test Records in a carrying case.)</p>

Publisher: Riverside Publishing Company
 425 Spring Lake Drive
 Itasca, IL 60143-2079
 Phone: 800-323-9540
 Website: www.riverpub.com

Strengths and Difficulties Questionnaire (SDQ)

SDQ:

-Domain: Socio-emotional

-Type of assessment: Screening

-Mode of assessment: Ratings/reports

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>A brief behavioral screening questionnaire, in several versions to meet the needs of researchers, clinicians and educationalists. Each version includes between one and three components: (1) 25 items on psychological attributes; (2) an impact supplement on the back to provide additional information to clinicians and researchers with an interest in psychiatric cases and the determinants of service use; and (3) follow-up questions for use after an intervention.</p> <p>3-16 years old</p>	<p>This instrument has been normed in the United States, Britain, Germany, Sweden, and Finland. In the U.S., the SDQ was included in the 2001 National Health Interview Survey Supplement.</p>	<p>For younger children, parents or teachers are respondents. Versions for adolescents are for self-completion.</p>	<p>The administrator should be familiar with the instrument and sensitive to the needs of the examinee. The tester should follow standard procedures, establish adequate rapport between the examiner and the examinee, and correctly score the examinee's responses.</p>	<p>Time needed not specified.</p> <p>Free and available in public domain in many languages.</p>

Publisher: Unpublished; developed by Robert Goodman in 1997.

Description and Questionnaire versions are available at <http://www.sdqinfo.com/b1.html>

Wechsler Intelligence Scales for Children (WISC)

WISC:

-Domain: Cognitive

-Type of assessment: Abilities

-Mode of assessment: Direct

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>Measures three IQs: Verbal Scale IQ, Performance Scale IQ, and Full Scale IQ</p> <p>6 -16 years</p>	<p>The normative sample is large (N=2,200) and representative of 1988 U.S. Census data. Separate norms were developed for translated versions. Available in Spanish, French, Italian, and Swedish. WISC-IV Spanish norms are calibrated to WISC-IV norms, which enables comparisons to English-speaking children. The norms have also been adjusted demographically to enable comparisons to children with similar U.S. educational experience and parental education level.</p>	<p>Child is respondent. Items include vocabulary similarities and comprehension. Performance block design and matrices and picture completion. Digit Span ask children to remember strings of numbers and repeat them back.</p>	<p>Training in administering WISC is necessary.</p>	<p>50-75 min</p> <p>Cost for entire kit (Administration and Scoring Manual, Integrated Technical and Interpretative Manual, Stimulus Book #1, 25 Record Forms, 25 Response Booklets #1, 25 Response Booklets #2, Symbol Search Scoring Key, Coding Scoring Key with Coding Recall, Cancellation Scoring Template, Block Design Cubes - 9) \$875</p>

Publisher: Harcourt Assessment
 19500 Bulverde Rd. San Antonio TX 78259.
 Phone: 1-800-211-8378
 Website: www.psychcorp.com

Wechsler Preschool and Primary Scales of intelligence (WPPSI)

WPPSI:

-Domain: Cognitive

-Type of assessment: Abilities

-Mode of assessment: Direct

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>A comprehensive assessment of general intellectual functioning. Can identify intellectual giftedness; cognitive intellectual delays; and mental retardation. Results can serve as a guide for placement decisions in clinical or school-related programs.</p> <p>2.5 years – 7y 3 months</p>	<p>The standardization sample included 1,700 children from across the U.S. The sample was stratified based on 2000 U.S. Census data for age, sex, race, parent education level, and geographic location.</p>	<p>The child is the respondent. Tasks vary by age and subtest. For verbal IQ tasks, children point to illustrations of objects, answer open-ended questions, translate outside knowledge into language, etc. Performance IQ tasks require children to work with puzzles, blocks, etc.</p>	<p>Administrators must have training and experience in the use of clinical instruments; e.g. experience with children whose characteristics are similar to those of the children being tested. Under some circumstances, trained teachers or examiners with supervision can administer this test.</p>	<p>Time need to complete test varies by age, ability, and motivation. It ranges from about 30 minutes to 60 minutes.</p> <p>Cost is \$850 (without carrying case) for all stimulus and manipulative materials, Examiner Manual, Technical Manual, 25 Record Forms for ages 2:6-3:11, 25 Record Forms for ages 4:0-7:3, and 25 Response Booklets.</p>

Publisher: Harcourt Assessment
 19500 Bulverde Rd. San Antonio TX 78259.
 Phone: 1-800-211-8378
 Website: www.psychcorp.com

Woodcock-Johnson (or Woodcock-Muñoz) (WJ)

WJ:
-Domain: Cognitive
-Type of assessment: Abilities
-Mode of assessment: Direct

Purpose and age range	Norms	Administration & Setting	Training needed for administration	Time needed and cost
<p>Determine an individual's cognitive strengths and weaknesses, to determine the nature of impairment, and to aid in diagnosis.</p> <p>2.5 years – adulthood</p>	<p>Nationally representative sample of 8,818 subjects drawn from 100 US communities using stratified random sampling. Also normed in Spanish.</p>	<p>Child is respondent. Complete certain tasks like pointing to a picture of a spoken word. Identifying two or three pieces that complete target shape.</p>	<p>Examiners should have thorough understanding of the administration and scoring procedure and formal training in assessment.</p>	<p>Each subtest 5-10 min, depending on number and combination of subtests. Cost for complete battery is \$966.50, plus an additional cost per individual tested.</p>

Publisher: Riverside Publishing
 425 Spring Lake Drive
 Itasca, IL 60143 Phone 1-800-323-9540
 Website: www.riverpub.com

APPENDIX B: New tests from developed countries

Australian Early Development Index (Australia)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Determine if child has the skills and developmental capacity to take advantage of the school's learning environment. Serves to predict later school outcomes. Continuous measure, but children scoring at the bottom 10% of each measure are considered "vulnerable." 4-5 years old	Teacher is respondent. Questions are relevant to the 5 domains, but also include child's special skills and special problems as well as demographic information.	Simple to administer with relatively little training	7 to 20 minutes. Sally A. Brinkman, sallyb@ichr.uwa.edu.au

Early Development Instrument (Canada)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Measures school readiness based on 5 domains: physical health and well-being, social competence, emotional maturity, language and communication, and cognitive development and general knowledge. Continuous measure.	Teacher is respondent. Questions are relevant to the 5 domains, but also include child's special skills and special problems as well as demographic information.	Administered as questionnaire that teachers complete on their own.	7 to 20 minutes. Dr. Magdalena Janus, Offord Center for Child Studies, McMaster University, 1200 Main St. W., Patterson Building Room 215, Hamilton, Ontario, Canada, L8N 3Z5

APPENDIX C: New tests from developing countries

Cambodian Developmental Assessment Test (Cambodia)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Measure level of cognitive, social, motor, and academic development for program evaluation based on country-specific standards	Child is tested with a series of items by a trained test administrator No norms available	Extensive training required	Not for sale, available from UNICEF Cambodia.

Early Childhood Care and Development Checklist (Philippines)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Monitors child's development in the following domains: fine and gross motor, receptive and expressive language, self-help, cognitive, and social-emotional. Continuous measure. 0-5 years	Ask child to complete various tasks that assess development in each of the domains. Normed for Filipino children. Sample of 10,915 children was drawn from urban and rural communities in Luzon, Visayas, and Mindanao.	Requires some experience with child development assessment.	Approximately 1.5 hours Available from authors

Escala Argentina de Inteligencia Sensorimotriz (EAIS)(Argentina)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Diagnostic qualitative measure of practical intelligence in the sensory-motor period based in the theories of Casati and Lezine, and Haessler. The test is based on the observation of the child's behavior in a variety of tasks (Piagetian). 6-24 months	Direct observation of children performing a series of tasks based on Piagetian tasks. No studies of sensitivity and specificity of the screening measures yet.	Unknown	Test developed by Oiberman et al.

Escala de Evaluación del Desarrollo Psicomotor (EEDP) (Chile)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Screening measure in the areas of language, social, coordination, and gross motor. Children are divided into three categories: normal, risk, and delayed. 0-24 months	Both observation and report. Normed for Chile.	Simple to administer with relatively little training	20 minutes Rodriquez, S. et al. (1996). Escala de evaluacion del desarrollo psicomotor: 0 a 24 meses. Santiago, Chile: 12th ed. Galdoc

Grover-Counter Scale of Cognitive Development, Revised (South Africa)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Measure of level of cognitive functioning (within defined range) of persons with impaired verbal skills, whether receptive, expressive, or both.	Child is asked to complete a variety of activities such as pattern completion, discrimination and grouping, and problem solving. Normed on 5 different samples of both children and mentally handicapped adults.	The tester should be a psychologist and should receive training on the administration of the test and the interpretation of the results.	Does not exceed 30 minutes. Less time for less able subjects because the more difficult sections will not be presented. Arvin Bhana, abhana@hsr.ac.za

Guide for Monitoring Child Development (Turkey)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Provide a method for developmental monitoring and early detection of developmental difficulties in children of low and middle income countries. 0-24 months	Caregiver is respondent. Complete a brief, open-ended, precoded interview. Questions pertain to child's social, emotional, and cognitive development.	Requires brief training (1 hour seminar and 1.5 hour practicum) and no background in child development	7 to 20 minutes. Ilgi Ertem, ertemilgi@yahoo.com

ICMR Psychosocial Developmental Screening Test (India)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
<p>Screens children for delays in five major developmental areas: (1) gross motor, (2) vision and fine motor, (3) hearing language and concept development, (4) personal skills, and (5) social skills. Test is continuous but may be used as a screening measure.</p> <p>0-6 years old</p>	<p>Observation and report</p> <p>Standardized on more than 13,000 children age 0 to 6 years in India</p>	<p>Must have educational qualifications similar to a community health care worker.</p>	<p>Shahnaz Vazir, s_vazir@hotmail.com</p>

IEA Preprimary Program Assessments (Multi-national)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
<p>The Child Activities (CA) system was used to document the activities and interactions of the target child in a given setting. The observations were coded into a predetermined category system that included 12 major categories. Conducted at age 4 years.</p> <p>Language and cognitive assessments at age 4 and 7 years.</p> <p>4-7 year olds.</p> <p>Used in 15 different countries (including Indonesia and Thailand).</p>	<p>Observation and direct testing.</p> <p>Norms not available. However, results were evaluated cross-nationally to determine that items assessed development similarly across all study sites.</p> <p>All tests were translated from English to the appropriate language(s).</p>	<p>A common set of procedures was used to train data collectors in all participating countries. The representatives who supervised the training and data collection in each country were first trained by researchers from the study's International Coordinating Center (ICC) and then trained data collectors in their own country. Data collectors were persons with experience in early childhood, such as teachers or graduate students in the field. Each one had to reach or exceed 80% agreement with a standard before being certified to collect data on each instrument.</p>	<p>40 minutes (two 20 minute sessions) for observations; no time cost determined for language and cognitive tests.</p> <p>Montie, J. E., Xiang, Z., & Schweinhart, L. J. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. <i>Early Childhood Research Quarterly</i>, 21, 313–331.</p> <p>Contact information: jmontie@highscope.org or lschweinhart@highscope.org</p> <p>Also see: http://www.highscope.org/file/Research/international/IEAInstruments/ChildActivitiesObsSystem.pdf for Child Activities Observation System.</p>

Kilifi Developmental Inventory (Kenya)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Measures psychomotor development, including locomotor skills and hand-eye coordination. KDI is a revised version of the KDC in order to include children under 1 year. It also eliminated the hearing, speech, language, and social emotional section. It is a continuous measure, but can be used for screening. 6-35 months	Child is asked to perform a motor activity after it is thoroughly described by an assessor. Dichotomous rating for each activity. Normed on a small group of children (roughly 200) from urban and rural populations. Authors created reference tables.	Preferably possesses a diploma in special education, or early childhood education, or equivalent. Training takes approximately 1 to 2 months.	Approximately 1 hour. Dr. Amina Abubakar, A.AbubakarAli@uvt.nl

Kilifi Developmental Checklist (Kenya)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Assesses psychomotor development in a resource-limited setting. Oringally designed to assess effects of malaria on functioning. Continuous measure. 1 to 5 years	Child is presented with various items that assess locomotor, hearing, speech, and language abilities. Ranked on a 1-3 scale.	Requires little training and little experience with child development assessment.	Dr. Amina Abubakar, A.AbubakarAli@uvt.nl

Parental Report Scales (Tanzania, Nepal)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Assesses child's language and motor abilities by asking questions to parents. Test is continuous. 6-59 months of age.	Parents are asked whether child can perform each behavior. Items are scored as Yes/No and scores (Yes responses) are summed.	Requires some training. Workers must be standardized.	Test takes 15 minutes. Scoring takes 5 minutes. Dr. Jane Kvalsvig kvalsvig@gmail.com or Dr. Patricia Kariger patriciakariger@gmail.com

Shoklo Developmental Test (Thailand)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
<p>Tests neurodevelopmental status of infants. Assessed motor, cognitive, social-emotional behaviors and speech. Purpose was to have an adaptable test with a low-resource setting and briefly trained testers.</p> <p>3-12 months</p>	<p>Observation and some parental report. Infants perform various activities using an assortment of toys and are scored on a pass/fail basis. Validated with Griffiths in the UK; no norms</p>	<p>Requires some training. Must be performed by a health care worker, but no experience related to child development required.</p>	<p>Test takes 20 minutes. Scoring takes 5 minutes.</p>

Shoklo Neurological Test (Thailand)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
<p>To evaluate abrupt neurological disturbances in children 9 to 36 months of age. Consists of three parts: assessment of coordination, tone, and behavior. Test is continuous. Purpose was to have an adaptable test with a low-resource setting and briefly trained testers.</p> <p>9-36 months</p>	<p>Observe infants while they perform various activities using an assortment of toys and are scored on a pass/fail basis. Tested with 300 Burmese infants in Thai camp. No clear age-based norms established.</p>	<p>Requires some training. Must be performed by a health care worker, but no experience related to child development required.</p>	<p>Test takes 15 minutes. Scoring takes 5 minutes.</p>

Test de Desarrollo Psicomotor (TEPSI) (Chile)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
<p>Evaluate child's development in three basic areas--motor function, coordination, and language--by observing child's behavior in certain situations set up by the examiner. Test is continuous, but may be used as a screening measure. (Standards for normal, at risk, and delayed children.)</p>	<p>Child is asked to perform various activities based on the area of development in question. Norms from a Chilean population</p>	<p>Simple to administer; minimum of training; done by preschool educators</p>	<p>30 to 40 minutes</p> <p>Hausler, IM & Marchant, T. (1980). TEPSI test de desarrollo psicomotor 2-5 anos (TEPSI test of psychomotor development 2-5 years). Ediciones Universidad Catolica, Octava Edicion 1999, Santiago, Chile.</p>

Yoruba Mental Subscale (Nigeria)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
<p>To measure mental abilities in young Yoruba children (Nigeria).</p> <p>Based on the Bayley Scales of Mental Development (Bayley, 1969). Includes 15 of 25 original items; those excluded did not perform well in analyses of piloting.</p> <p>Items were related to parent ratings of responsibility for carrying out certain tasks.</p> <p>22–26 months old</p>	<p>Home; norms not established.</p>	<p>University students or graduates of psychology or related disciplines.</p> <p>Time for training not specified, but should be similar to that required for administering the Bayley (1-2 months).</p> <p>Standardization of administrators required.</p>	<p>Time not specified. Based on the number of items, this test should take about 30-45 minutes to complete.</p> <p>Ogunnaike, O.A., & Houser Jr., R.F. (2002). Yoruba toddlers' engagement in errands and cognitive performance on the Yoruba Mental Subscale. <i>International Journal of Behavioral Development</i>, 26(2), 145-153.</p>

APPENDIX D: Details of tests developed to measure executive function

Bayley Examiner Assessment

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Observer reports of children's self-regulatory abilities during a cognitive testing situation. Birth to 42 months	Administered as a part of assessments of cognitive development	Some training needed for observers to be reliable with one another	5-10 minutes Purchased as part of Bayley package. Harcourt Assessment

Backward Digit Span Test

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Assesses working memory and inhibition. Children are instructed to repeat 2-6 digits backwards. ~3-6 years	Direct; child is respondent. Can be done at home or in clinical/school setting. Requires some materials.	Some training required. Prefer minimum of high school education.	Not stated --15 minutes? Reference: Davis, H. L., & Pratt, C. (1996). The development of children's theory of mind: The working memory explanation. <i>Australian Journal of Psychology</i> , 47, 25-31.

Behavioral Rating Inventory of Executive Function-P (BRIEF-P)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
The BRIEF-P is a standardized rating scale of 63 items to measure behavioral manifestations of executive function in preschool children. 2-5 years	Report; used by parents, teachers, and day care providers to rate a child's executive functions within the context of his or her everyday environments--both home and preschool. Normative data are based on representative US sample	Some training required. Prefer minimum of high school education.	10-15 minutes \$150.00 for complete kit: Includes Manual; 25 Rating Forms; 25 Scoring Summary/Profile Forms Contact: Western Psychological Services 12031 Wilshire Blvd. Los Angeles, CA 90025-1251 Telephone: (800) 648-8857 - FAX: (310) 478-7838

Delay of Gratification Test

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Measures children's abilities to delay a short-term reward in order to gain a larger, longer-term reward. Delay of gratification has been linked to children's performance in school, beyond contributions of IQ.	Child is presented with a choice between a short-term reward or a larger reward that is only given if the child waits a certain amount of time (usually no more than 3 minutes for young children). No norms available	Requires some training of administrator.	5 minutes. No cost for administering procedure other than rewards for children. Several versions of the test; can be developed in country.

Leiter Examiner Report

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Observer reports of children's self-regulatory abilities during a cognitive testing situation. Birth through childhood	Administered as a part of assessments of cognitive development No norms available	Some training needed for observers to be reliable with one another	5-10 minutes Purchased as part of Leiter package Psychological Assessment Resources

Stroop test (adapted for younger children as Day/Night test)

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Assesses working memory and inhibition in a Stroop-like task. Children are instructed to say "night" when shown a card with a sun, and "day" when shown a card with moon and stars. 3-6 years	Direct; child is respondent. Can be done at home or in clinical/school setting. Requires some materials.	Some training required. Prefer minimum of high school education.	Not stated --15 minutes? See reference: Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3.5–7 years old on a Stroop-like day-night test. <i>Cognition</i> , 53, 129–153. Contact: diamond@catell.psych.upenn.edu

Wisconsin card-sorting task

Purpose and age range	Administration Setting & norms	Training needed for administration	Time needed and cost / Author of test
Measures self-control. Assesses children's abilities to inhibit previously-learned responses when presented with a new rule. 4-6 years	Task involves presenting children with cards that must be sorted according to one dimension (such as color) and then, in a second game, by another dimension (such as shape). Children are scored based on their ability to inhibit the response learned in one game when playing the second.	Requires some training	10 to 15 minute to administer, 10 minutes to score. Cost for manual and administration procedures is about \$250. Available for purchase from Psychological Assessment Resources.

APPENDIX E: Table of measures used and where

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Achenbach Child Behavior Checklist	N/A	Brazil	4-5 years old (older than 5 included in the study as well)	Not specified in abstract or methods	"Six hundred and one children randomly selected from a Brazilian birth cohort were evaluated for behavioral/emotional problems through mother interview at 4 and 12 years with the same standard procedure - Child Behavior Checklist (CBCL)."	"CBCL Total Problem score presented a medium stability ($r = .42$) with externalizing problems showing higher stability and more homotypic continuity than internalizing problems. Of the children presenting deviant scores at the age of 4, only 31% remained deviant at the age of 12 ($p < .001$). A deviant CBCL Total Problem score at 12 years old was predicted by Rule-Breaking Behavior [OR = 7.46, 95% CI 2.76-20.19] and Social Problems [OR = 3.56, 95% CI 1.36-9.30] scores at 4 years of age. Either Rule-Breaking or Aggressive Behavior - externalizing syndromes - were part of the predictors for the three broad-band CBCL scores and six out of the eight CBCL syndromes. Behavioral/emotional problems in preschool children persist moderately up to pre-adolescence in a community sample. Externalizing problems at the age of 4 comprise the developmental history of most behavioral/emotional problems at pre-adolescence. Our findings concur with findings from developed countries and are quite similar for continuity, stability and predictability."	Anselmi L, Barros FC, Teodoro ML, Piccinini CA, Menezes AM, Araujo CL, Rohde LA.	2008	Anselmi L, Barros FC, Teodoro ML, Piccinini CA, Menezes AM, Araujo CL, Rohde LA. Continuity of behavioral and emotional problems from pre-school years to pre-adolescence in a developing country. <i>J Child Psychol Psychiatry</i> . 2008 May;49(5):499-507. Epub 2008 Mar 10. <u>Other child development tests used:</u> N/A
Achenbach Child Behavior Checklist	N/A	China	children, age not quantified in abstract	Not specified in abstract. CBCL edited by XU Tao-yuan in 1992	"The present study was designed to investigate the influence of Congenital Heart Disease (CHD) on the mentality and behavior in children, and to compare post operative mentality and behavior in children receiving interventional therapy and congenital heart surgery."	"The mentality and behavior abnormal rates of the boys and girls suffering from CHD were significantly higher than those of controls ($P < 0.01$, $P < 0.05$). The behavior abnormalities of the boys presented as depression, social flinch, physical complains, assault and violate rules. Whereas the girls presented as depression, social flinch, physical complains and violate rules. The total cursory mark of postoperative check result of the interventional and surgical children, both in girls and in boys, were significantly lower than those of the preoperative children ($P < 0.05$). The total and assault cursory mark of postoperative check result of children treated with interventional therapy were significantly lower than those of children treated with the surgical operations ($P < 0.05$). The abnormal rates of mentality and behavior positively correlated with the disease course. CHD is associated with increased abnormal mentality and behavior of the children. Early treatment, especially the interventional therapy can significantly improve the mentality and behavior of the children with CHD."	Zhang K, Wang YB, Li YP, Liu F, Zhang ZH, Wang ZX, Hao FZ.	2008	Zhang K, Wang YB, Li YP, Liu F, Zhang ZH, Wang ZX, Hao FZ. [Mentality and behavior of children with congenital heart diseases] <i>Zhonghua Xin Xue Guan Bing Za Zhi</i> . 2008 May;36(5):418-21 <u>Other child development tests used:</u> N/A

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Achenbach Child Behavior Checklist	N/A	China	children 5 years old (older than 5 included in the study as well)	Not specified in abstract or methods	"This study sought to determine the prevalence of mood disorders among patients with microtia and to explore clinical features associated with mood disorders. "	"The prevalence of mood disorders among microtia patients: 'depression' 20.2%, 'interpersonal sensitivity/social difficulties' 36.6% and 'hostility/aggression' 26.3%. Multivariate analyses suggested that age of patients, severity of microtia, low levels of maternal education, being teased by peers, family disharmony, psychological impact on parents and overprotection from parents are significantly associated with mood disorders of patients. Our findings suggest that microtia patients exhibit three significant mood disorders including depression, interpersonal sensitivity/social difficulties and hostility/aggression. Some risk factors should be actively prevented and controlled, such as being teased by peers, family disharmony, psychological impact on parents and overprotection from family."	Jiamei D, Jiake C, Hongxing Z, Wanhou G, Yan W, Gaifen L.	2008	Jiamei D, Jiake C, Hongxing Z, Wanhou G, Yan W, Gaifen L. An investigation of psychological profiles and risk factors in congenital microtia patients. J Plast Reconstr Aesthet Surg. 2008;61 Suppl 1:S37-43. Epub 2007 Nov 5. <u>Other child development tests used:</u> Symptom Checklist-90
Achenbach Child Behavior Checklist	N/A	China	children 3-5 years old	Chinese	"This study aims to test the hypothesis that children who have a regular FP (family physician) have better health behaviours and psychosocial health than children who do not have a regular FP. "	"Good health and hygiene behaviours were significantly more prevalent in children who had a regular FP. Children who did not have a regular FP had statistically significant higher scores in all three main domains of the CBCL. Children with a regular FP had higher odds ratios for various hygiene and health behaviours after adjusting for socioeconomic status. The findings highlight the potential role of FPs in promoting health, hygiene, and wellbeing in children."	Lee A, Wong WS, Fung WY, Leung PW, Lam C.	2007	Lee A, Wong WS, Fung WY, Leung PW, Lam C. Children with a regular FP - do they have better health behaviours and psychosocial health? Aust Fam Physician. 2007 Mar;36(3):180-2, 191. <u>Other child development tests used:</u> N/A
Achenbach Child Behavior Checklist	N/A	China	children 4-5 years old	Not specified in abstract	"This study investigated whether breastfeeding is associated with the occurrence of behavioral problems and the temperament development in preschool children."	"After controlling for confounding variables, such as family income and parental education levels, it was found that a breastfeeding duration of ≥ 9 months was a protective factor against behavioral problem occurrence in boys (OR=0.184). In girls, a breastfeeding duration of ≥ 9 months was also a protective factor against behavioral problem occurrence (OR=0.165), while a mixed feeding with more breast milk and less formula milk was a risk factor (OR=2.203). The factors influencing temperament development consisted of exclusive formula feeding and the duration of breastfeeding (lasting for 4-6 months or 7-9 months) as well as a mixed feeding (with more formula milk and less breast milk, more breast milk and less formula milk, or equal amount of both). The fewer amounts and the shorter duration of breastfeeding are risk factors for behavioral problems occurrence in children aged 4-5 years. Children's temperament development is correlated with the feeding patterns and the breastfeeding duration"	Liu F, Ma LJ, Yi MJ	2006	Liu F, Ma LJ, Yi MJ. [Association of breastfeeding with behavioral problems and temperament development in children aged 4-5 years] Zhongguo Dang Dai Er Ke Za Zhi. 2006 Aug;8(4):334-7. <u>Other child development tests used:</u> temperment quetionnaire and self-designed inventory quetionnaire

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Achenbach Child Behavior Checklist	N/A	China	children, age not quantified in abstract	Not specified in abstract	"To collect basic information on family burdens and long-term influence of children suffered from traumatic brain injury (TBI)."	"The mean adaptation partnership growth affection and resolve scale (APGAR) score of 113 children before TBI was 7.96 and score after TBI was 6.94, which had significantly difference through t test. The mean APGAR score after 6 months was 7.60, which was significantly different from the hospital data. Among group with severe TBI, the family APGAR score in hospital was significantly smaller than that before injury occurred, and the family APGAR score in 6 months after being discharged from the hospital had no significant difference with the score when staying in the hospital. The three leading dimensions among family burden scale of diseases (FBS) scores after TBI were dimension of family economic burden, family daily life and family entertainment. 6 months later, the three leading dimensions had changed to be as dimension of mental health status, dimension of family relationship and dimension of family economic burden. Mean score of child behavior checklist (CBCL) assessed at 6-months follow up period among 113 children was among normal range. Family function of children with TBI was affected by TBI. However, family function could be recovered along with child's convalescence except among children with severe TBI. Long-term pressure of TBI on family was revealed in mental health status and family relationship. In this study, there were no evidence of association between TBI and children's behavior problem."	Chen H, Meng H, Lu ZX.	2006	Chen H, Meng H, Lu ZX. [Prospective study on family burden following traumatic brain injury in children] Zhonghua Liu Xing Bing Xue Za Zhi. 2006 Apr;27(4):307-10. <u>Other child development tests used:</u> adaptation partnership growth affection and resolve scale (APGAR)
Achenbach Child Behavior Checklist	N/A	China	children, age not quantified in abstract	Not specified in abstract. CBCL edited by Gong Yao-xian in 1986	"The present study was designed to investigate the influence of viral myocarditis on mental behavior of the children."	"The mentality and behavior of the children with viral myocarditis were distinctly abnormal. The abnormal rates of boys and girls suffering from acute and deferment viral myocarditis were evidently higher than those of control children (P < 0.01). The behavioral abnormalities of boys were hypochondria, social difficulties, unwell of body and attack. Whereas, the girls presented hypochondria, unwell of body, social flinch and disobeyed discipline, which was significantly different from the control children. The total and hypochondria cursory mark of the second check result of deferment boys were evidently higher than those of the first check (P < 0.05). The total cursory mark of the second check result of deferment girls was higher than that of the first check (P < 0.05) and evidently higher than that of the acute second check result (P < 0.01). The abnormal rates of mentality and behavior correlated positively with the age of children and they were associated with	Wang ZX, Xu L, Wang YL, Zhang KX, Zhang K, Zhang ZH.		Wang ZX, Xu L, Wang YL, Zhang KX, Zhang K, Zhang ZH. [Mentality and behavior of children suffering from viral myocarditis] Zhonghua Er Ke Za Zhi. 2006 Feb;44(2):122-5. <u>Other child development tests used:</u> N/A

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						the severity of the illness. Viral myocarditis evidently affected the mentality and behavior of children, which should be paid great attention to."			
Achenbach Child Behavior Checklist	N/A	Dominican Republic	children 2-5 years old (older than 5 included in the study as well)	Not specified in abstract	"The purpose of this report is to describe behavioural problems encountered in a group of Dominican children living with Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome (HIV/AIDS) in the Dominican Republic. They were not receiving antiretroviral treatment. "	"Descriptive statistics revealed a high proportion of the children, both younger (approximately 40%) and older (46%) scored in the borderline/clinical ranges for internalizing problems, including anxiety, withdrawn-depressed and somatic complaints. In addition, 46% of the older children were perceived as having externalizing problems (rule breaking and aggressive behaviour). These findings suggest that a high incidence of behavioural and mood problems may be prevalent among Dominican children with HIV The findings are discussed in terms of future research to examine other risk factors that might contribute to the high rate of maladaptive behaviours observed in the present report, including the contribution of socio-economic status, caregiver illness, caregiver education and parental loss."	Mendoza R, Hernandez-Reif M, Castillo R, Burgos N, Zhang G, Shor-Posner G.	2007	Mendoza R, Hernandez-Reif M, Castillo R, Burgos N, Zhang G, Shor-Posner G. Behavioural symptoms of children with HIV infection living in the Dominican Republic. West Indian Med J. 2007 Jan;56(1):55-9. <u>Other child development tests used:</u> N/A
Achenbach Child Behavior Checklist	N/A	India	4 to 11 years	Not specified	Study of the extent and nature of psychiatric disorders in school children in a defined geographical area and their psychosocial correlates. The parent interview (Stage II) for all children on the Childhood Psychopathology Measurement Schedule was used, an Indian adaptation of Achenbach's CBCL.	6.33 per cent of the children studied (n = 963) were found to have psychiatric disorders on ICD-10 criteria. Teacher's estimation of the prevalence rates was higher, i.e., 10.17 per cent as compared to parent's estimate i.e., 7.48 per cent. The most prevalent disorder was enuresis.	Malhotra et al.	2002	Malhotra S, Kohli A, Arun P. Prevalence of psychiatric disorders in school children in Chandigarh, India. Indian J Med Res. Jul 2002;116:21-28.
Achenbach Child Behavior Checklist	N/A	Iraqi Kurdistan	4 to 16 years	Kurdish	Two-year follow up of study comparing competency scores, behavioral problems, and PTSD of orphans in foster care and orphans in modern orphanages.	Although both samples revealed significant decrease in the means of total competence and problem scores over time, the improvement in activity scale, externalizing problem scores and post-traumatic stress disorder-related symptoms proved to be more significant in the foster care than in the orphanages. While the activity scale improved in the foster care, the school competence deteriorated in both samples, particularly among the girls in the orphanages.	Ahmad et al.	2005	Ahmad A, Qahar J, Siddiq A, et al. A 2-year follow-up of orphans' competence, socioemotional problems and post-traumatic stress symptoms in traditional foster care and orphanages in Iraqi Kurdistan. Child Care Health Dev. Mar 2005;31(2):203-215.
Achenbach Child Behavior Checklist	N/A	Iraqi Kurdistan	4 to 16 years	Kurdish	Study comparing competency scores, behavioral problems, and PTSD of orphans in foster care and orphans in modern orphanages, competency scores and behavioral	While competency scores showed an improvement in both samples at the follow-up test, the problem scores increased in the orphanage sample and decreased among the foster care subjects. Moreover, the orphanage sample reported higher frequency of post-traumatic stress disorder (PTSD) than the foster care children.	Ahmad et al.	1996	Ahmad A, Mohamad K. The socioemotional development of orphans in orphanages and traditional foster care in Iraqi Kurdistan. Child Abuse Negl. Dec 1996;20(12):1161-1173.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					problems were assessed at baseline and at a 1-year follow-up. PTSD reactions were examined at a 1-year follow-up.				
Achenbach Child Behavior Checklist	N/A	Israel	6 to 11	Hebrew	Child Behavior Checklists (CBCL) were completed by parents of 450 clinically referred Israeli boys aged 6-11.	The first seven of ten specific syndromes were highly correlated with American and Dutch syndromes derived from the CBCL providing further evidence of their cross-cultural robustness.	Auerbach et al.	1991	Auerbach JG, Lerner Y. Syndromes derived from the Child Behavior Checklist for clinically referred Israeli boys aged 6-11: a research note. <i>J Child Psychol Psychiatry</i> . Sep 1991;32(6):1017-1024.
Achenbach Child Behavior Checklist	N/A	Israel	11 to 16 years	Hebrew	Study on the extent of behavior problems in Israeli adolescents suffering from chronic illness. A comparison was made between parent-reported and self-reported behavioral symptomatology using the CBCL and another measure. Children suffering from cystic fibrosis, asthma, or hematological/oncological conditions were assessed.	Parent- and self-reports were significantly positively correlated in the group of chronically ill children and two comparison groups (all chronically ill children $r = .22$; Healthy group $r = .27$; psychiatric group $r = .50$), but the correlations were particularly low (and non-significant) in younger adolescents with hematological/oncological conditions or HCF., pointing to the need for physicians to include parents' and adolescents' viewpoints in their assessments of these adolescents' psychosocial state. The mean number of parent-reported and self-reported behavior problems in the illness groups was no different from that of the Healthy group	Stawski et al.	1995	Stawski M, Auerbach JG, Barasch M, Lerner Y, Zimin R, Miller MS. Behavioral problems of adolescents with chronic physical illness: a comparison of parent-report and self-report measures. <i>Eur Child Adolesc Psychiatry</i> . Jan 1995;4(1):14-20.
Achenbach Child Behavior Checklist	N/A	Malaysia	4 to 12 years	Not specified	Study to compare parenting stress among Malaysian mothers of children with mental retardation and a control group, and to determine factors associated with stress.	The total child behaviour scores from the CBCL ($P < 0.01$), IQ scores ($P < 0.01$) and sibship size ($P < 0.01$) were associated with child-related domain scores. A large proportion of mothers of children with mental retardation experienced substantial parenting stress, especially Chinese and unemployed mothers, and this warrants appropriate intervention.	Ong et al.	1999	Ong L, Chandran V, Peng R. Stress experienced by mothers of Malaysian children with mental retardation. <i>J Paediatr Child Health</i> . Aug 1999;35(4):358-362.
Achenbach Child Behavior Checklist	N/A	Netherlands	7 years	Not specified	Study to examine the development and adjustment of 7-year-old children adopted in infancy.	The study provides evidence of an increased risk for behavior problems of infant-placed 7-year-old internationally, transracially adopted children in the Netherlands. However, parents reported more behavior problems for adopted boys than for adopted girls. Notably, about 30% of the adopted children were classified as clinical on the CBCL scale for total problems, which is a much larger percentage than the 10% found in the normative population. It was suggested that these results could be explained by the operation of multiple risk factors before and after adoption placement, e.g. the child's genetic disposition, pre-natal and pre-adoption care, or the child's cognitive understanding of adoption in middle childhood. Also, results	Stams et al.	2004	Stams GJ, Juffer F, Rispen J, Hoksbergen RA. The development and adjustment of 7-year-old children adopted in infancy. <i>J Child Psychol Psychiatry</i> . Nov 2000;41(8):1025-1037.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						suggest that maternal sensitive responsiveness in adoptive families declines in the transition from early to middle childhood.			
Achenbach Child Behavior Checklist	N/A	Netherlands	11 to 18 years	Moroccan-Arabic	Study to examine factors associated with internalizing problems in immigrant Moroccan youth.	The data showed relations between internalizing problems and several child (externalizing and chronic health problems), proximal family (paternal and maternal support and parent-child conflict), contextual family (conflicts between parents about parenting and total number of life-events), school/peer (being bored), and migration variables (adolescent's perceived discrimination). Moreover, a modest relation was found between internalizing problems and parental psychopathology. Few associations occurred with family educational level.	Stevens et al (14)	2005	Stevens GW, Vollebergh WA, Pels TV, Crijnen AA. Predicting internalizing problems in Moroccan immigrant adolescents in The Netherlands. Soc Psychiatry Psychiatr Epidemiol. Dec 2005;40(12):1003-1011.
Achenbach Child Behavior Checklist	N/A	Netherlands	11 to 18 years	Moroccan-Arabic	Study of the predictors of externalizing problems in Moroccan immigrant adolescents.	There was a clear association between externalizing problems and several factors; child (gender, internalizing problems), proximal family (parental monitoring and affection, support from father and mother, and parent-child conflict), contextual family (conflicts between parents about parenting, destructive communication between parents, and total number of life-events), school/peer (problems at school, involvement with deviant peers, hanging out), and migration variables (adolescent's perceived discrimination). Hardly any association was observed between externalizing problems and parental psychopathology, and between externalizing problems and family employment level. Most findings matched results found in earlier studies on non-immigrant youth.	Stevens et al (14)	2005	Stevens GW, Vollebergh WA, Pels TV, Crijnen AA. Predicting externalizing problems in Moroccan immigrant adolescents in the Netherlands. Soc Psychiatry Psychiatr Epidemiol. Jul 2005;40(7):571-579.
Achenbach Child Behavior Checklist	N/A	Netherlands	14 to 18 years	Moroccan-Arabic, Turkish, and Dutch	Study comparing emotional and behavioral problems of Moroccan immigrant children to those of Dutch native children and Turkish immigrant children.	Moroccan parents reported as many problems as Dutch parents, but less problems than Turkish parents. Teachers, however, reported substantially more externalizing problems for Moroccan pupils compared to Dutch and Turkish pupils. Moroccan adolescents themselves reported less problems than Dutch and Turkish adolescents.	Stevens et al.	2003	Stevens GW, Pels T, Bengi-Arslan L, Verhulst FC, Vollebergh WA, Crijnen AA. Parent, teacher and self-reported problem behavior in The Netherlands: comparing Moroccan immigrant with Dutch and with Turkish immigrant children and adolescents. Soc Psychiatry Psychiatr Epidemiol. Oct 2003;38(10):576-585
Achenbach Child Behavior Checklist	S	Sri Lanka	children 5 years old (older than 5 included in the study as well)	Sinhala translation (utilizing translation and back-translation)	"To translate the child behaviour checklist (CBCL) into Sinhala and validate it for assessment of mental health status of children aged 5-10 years."	"Semantics, content, and conceptual and criterion validity of CBCL-S were satisfactory. At the cut-off level of 39, CBCL-S had a sensitivity of 90% and a specificity of 88% for boys and a sensitivity of 89% and a specificity of 92% for girls. Internal consistency, test-retest reliability, and inter-	Senaratna BC, Perera H, Fonseka P.	2008	Senaratna BC, Perera H, Fonseka P. Sinhala translation of child behaviour checklist: validity and reliability. Ceylon Med J. 2008 Jun;53(2):40-4.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
				and validation		interviewer reliability of CBCL-S were satisfactory. CBCL-S is a valid and reliable instrument to measure mental health status of Sinhalese children aged 5-10 years in Sri Lanka."			<u>Other child development tests used:</u> N/A
Achenbach Child Behavior Checklist	N/A	Sweden	6 to 18 years	Not specified	Study to assess types and scores of traumatic experiences, post-traumatic stress symptom and behavioural disorders among Kurdistanian refugee children in Sweden and a comparable group of Swedish children.	No significant differences were found between the 2 samples regarding types of traumatic events, frequencies of post-traumatic stress disorder, post-traumatic stress symptom scores or behavioural problem scores, except in 3 aspects: Kurdistanian children reported more war experience and being lost, while Swedish children presented higher frequencies of leisure-time accidents.	Sundelin et al	2001	Sundelin Wahlsten V, Ahmad A, von Knorring AL. Traumatic experiences and posttraumatic stress reactions in children and their parents from Kurdistan and Sweden. Nord J Psychiatry. 2001;55(6):395-400.
Achenbach Child Behavior Checklist	N/A	Taiwan	children, age not quantified in abstract	Not specified in abstract	"In this study, we used the Child Behavior Checklist (CBCL) to determine a behavioral profile for children with chronic epilepsy."	"We found behavioral disturbances in 42% (n=24) of the epileptic patients and in 8% (n=4) of the controls. No significant differences were found between patients with and without behavioral problems on the clinical variables. Behavioral problems deserve special attention in children with epilepsy. CBCL can be used as a screening instrument with these children."	Fang PC, Chen YJ.	2007	Fang PC, Chen YJ. Using the child behavior checklist to evaluate behavioral problems in children with epilepsy. Acta Paediatr Taiwan. 2007 Jul-Aug;48(4):181-5. <u>Other child development tests used:</u> N/A
Achenbach Child Behavior Checklist	N/A	Taiwan	children, age not quantified in abstract	Not specified in abstract	"In the current study, the behavior and emotional problems of 1,042 disabled children in special education programs were evaluated using the Chinese version of the Child Behavior Checklist (CBCL-C) and the Teacher's Report Form (TRF). "	"Using the 60th percentile on the two tests as a cutoff representing a clinical indication, students who reached this cutoff point but did not receive mental health services in the past six months were considered to have "unmet mental health needs." Of the special education students in the study 73.9% reached clinical indications, but did not receive mental health care."	Liang HY, Chang HL.	2007	Liang HY, Chang HL. Disabled children in special education programs in Taiwan: use of mental health services and unmet needs. Psychol Rep. 2007 Jun;100(3 Pt 1):915-23. <u>Other child development tests used:</u> N/A
Achenbach Child Behavior Checklist	N/A	Taiwan	children 5 years old (older than 5 included in the study as well)	Chinese	"Some evidence has indicated the TOVA can be useful in diagnosing ADHD. This study examines its validity and reliability in helping diagnose Taiwanese ADHD children. "	"Results showed a mean internal consistency of 0.81 for all six TOVA variables across conditions, with moderate convergent and discriminant validities. Groups showed significant differences in response time variability, D' and ADHD scores, with the normal group outperforming the ADHD group. Significant group differences were also found in all CBCL subscale scores except somatic complaints. The ADHD group obtained a clinically significant score on the hyperactivity subscale of the CBCL. The findings partially support the usefulness of the TOVA in assessing attention and impulsivity problems for a Taiwanese sample. Future studies should increase the sample size, use multiple measures, and collect behavior ratings from both	Wu YY, Huang YS, Chen YY, Chen CK, Chang TC, Chao CC.	2007	Wu YY, Huang YS, Chen YY, Chen CK, Chang TC, Chao CC. Psychometric study of the test of variables of attention: preliminary findings on Taiwanese children with attention-deficit/hyperactivity disorder. Psychiatry Clin Neurosci. 2007 Jun;61(3):211-8. <u>Other child development tests used:</u> Test of Variables of attention (TOVA)

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						parents and teachers."			
Achenbach Child Behavior Checklist	N/A	Taiwan	children 4-5 years old (older than 5 included in the study as well)	Chinese version, validated by Huang et al. 1994	"The aim of this study was to examine the effect of age, gender and perinatal risk factors on the risks for sleep problems, and investigate the relation between childhood sleep problems and children's behavioral syndromes and parental mental distress in early and middle childhood."	"Results showed that boys suffered from more sleep problems than girls. Early insomnia, sleep terrors and enuresis decreased with ages, but sleepwalking increased with ages. Perinatal exposure to alcohol, coffee and non-prescribed medication, vaginal bleeding, artificial delivery, first-born order and higher parental CHQ score (> or =4) were significantly associated with several childhood sleep problems. In addition, children with sleep problems had higher T-scores of the eight behavioral syndromes derived from the CBCL. Our findings indicated that the childhood sleep problems were associated with perinatal risk factors, parental psychopathology and children's behavioral problems."	Shang CY, Gau SS, Soong WT.	2006	Shang CY, Gau SS, Soong WT. Association between childhood sleep problems and perinatal factors, parental mental distress and behavioral problems. J Sleep Res. 2006 Mar;15(1):63-73. <u>Other child development tests used:</u> Chinese Health Questionnaire (CHQ)
Achenbach Child Behavior Checklist	N/A	Turkey	Adolescents, mean age 13.8	Turkish	Study to explore the type and frequency of psychopathology in a clinical as well as a non-clinical sample of obese adolescents, and in a normal weight control group.	The mean scores of anxiety-depression, social problems, social withdrawal and total problem in the CBCL scale of the clinical obese group were significantly higher than the non-clinical obese group and the normal weight control group. The results support previously published reports which show a higher ratio of psychopathology (depression, behavioral problems, low-esteem) among clinical obese adolescents than among non-clinical obese adolescents.	Ereemris et al.	2004	Ereemris S, Cetin N, Tamar M, Bukusoglu N, Akdeniz F, Goksen D. Is obesity a risk factor for psychopathology among adolescents? Pediatr Int. Jun 2004;46(3):296-301.
Achenbach Child Behavior Checklist	N/A	Turkey	Children and adolescents, age 5 to 18	Turkish	Study to evaluate the effects of internal displacement and resettlement within Turkey on the emotional and behavioral profile of children.	The children and adolescents with internal displacement had significantly higher internalizing, externalizing and total problem scores on the CBCL and other measures. The effect of displacement was related to higher internalizing problems when factors like physical illness, child age, child gender and urban residence were accounted. The overall effect was small explaining only 0.1-1.5% of the total variance by parent reports, and not evident by teacher reports. The results are consistent with previous immigration studies: child age, gender, presence of physical illness and urban residence were more important predictors of internalization and externalization problem scores irrespective of informant source.	Erol et al.	2005	Erol N, Simsek Z, Oner O, Munir K. Effects of internal displacement and resettlement on the mental health of Turkish children and adolescents. Eur Psychiatry. Mar 2005;20(2):152-157.
Achenbach Child Behavior Checklist	N/A	Turkey	Children and adolescents, age 4 to 18	Dutch	Study to compare self-reported emotional and behavioral problems for Turkish immigrant, native Dutch and native Turkish adolescents.	Turkish immigrant adolescents reported more problems in comparison to their Dutch and native Turkish peers. Different patterns of parent-child interaction, family values and delay of Dutch language skills are considered to be responsible for these differences in scores.	Janssen et al.	2004	Janssen MM, Verhulst FC, Bengi-Arslan L, Erol N, Salter CJ, Crijnen AA. Comparison of self-reported emotional and behavioral problems in Turkish immigrant, Dutch and Turkish adolescents. Soc Psychiatry

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
									Psychiatr Epidemiol. Feb 2004;39(2):133-140.
Achenbach Child Behavior Checklist	N/A	Turkey	grades 2 and 3 (mean age 7.95 years)	Turkish	Study to examine the competency and problem behavior correlates of television viewing in school-aged children using the CBCL.	Stepwise logistic regression analysis revealed that the only significant variables associated with a risk of watching television for more than 2 hours were age, gender, social subscale, and attention problem subscale scores of the CBCL. As evaluated by the CBCL, television viewing time is positively associated with social problems, delinquent behavior, aggressive behavior, externalization, and total problem scores. Older age, male gender, and decreasing social subscale and increasing attention problem subscale scores on the CBCL increases the risk of watching television for more than 2 hours.	Ozmert et al.	2002	Ozmert E, Toyran M, Yurdakok K. Behavioral correlates of television viewing in primary school children evaluated by the child behavior checklist. Arch Pediatr Adolesc Med. Sep 2002;156(9):910-914.
Achenbach Child Behavior Checklist	Version 2/3	Turkey	14-43 months old	Turkish translation and validation	"We investigated the congruent and criterion validity of the Aberrant Behavior Checklist (ABC) in a clinical sample of toddlers seen over 1 year in Turkey."	"The total ABC score, which is interdependent with subscales (e.g., Irritability, Social Withdrawal) of the ABC, was significantly correlated with the CBCL-total ($r = .73$) and AuBC-total ($r = .71$) scores. Subscales of the ABC revealed significant differences between diagnostic groups. ABC Total, and the Irritability and Hyperactivity subscale scores, were significantly higher in children with externalizing disorders; the Lethargy/Social Withdrawal and Stereotypic Behavior subscale scores were significantly higher in toddlers with autism. The ABC appears to be capable of discriminating several syndromes, such as disruptive behavior disorders and autism in early childhood."	Karabekiroglu K, Aman MG.	2008	Karabekiroglu K, Aman MG. Validity of the aberrant behavior checklist in a clinical sample of toddlers. Child Psychiatry Hum Dev. 2009 Mar;40(1):99-110. Epub 2008 Jul 4. <u>Other child development tests used:</u> Autism Behavior Checklist (AuBC)
Achenbach Child Behavior Checklist	N/A	Turkey	children 5 years old (older than 5 included in the study as well)	Not specified in abstract	"The aim of this study was to assess the effects of a 14-week swimming training program on the competence, problem behaviour, and body awareness in 13 children with cerebral palsy aged 5 to 10 years, compared with 10 subjects in a comparison group."	"The results showed that swimming training produced significant gain on body awareness in the Swimming Group, whereas no significant group differences were evident in competence and problem behaviours on parent or teacher forms of the CBCL."	Ozer D, Nalbant S, Aktop A, Duman O, Keleş I, Toraman NF.	2007	Ozer D, Nalbant S, Aktop A, Duman O, Keleş I, Toraman NF. Swimming training program for children with cerebral palsy: body perceptions, problem behaviour, and competence. Percept Mot Skills. 2007 Dec;105(3 Pt 1):777-87. <u>Other child development tests used:</u> Body Awareness
Achenbach Child Behavior Checklist	N/A	Turkey	children 4-5 years old (older than 5 included in the study as well)	Not specified in abstract	"To evaluate the epidemiology of attention problems using parent, teacher, and youth informants among a nationally representative	"The CBCL and TRF attention problems scores were higher among young male children, whereas the YSR reported scores were higher among older adolescents without a gender effect. The CBCL and YSR scores were also higher by urban residence. Compared with other European samples, our	Erol N, Simsek Z, Oner O, Munir K.	2008	Erol N, Simsek Z, Oner O, Munir K. Epidemiology of attention problems among Turkish children and adolescents: a national study. J Atten Disord. 2008

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					Turkish sample."	national sample had higher mean attention problems scores than the Scandinavian but lower mean scores than the former Soviet Union samples. In addition to elucidating the profile of attention problems in Turkey, our results also contribute to understanding the comparative global epidemiology of attention problems."			Mar;11(5):538-45. Epub 2008 Jan 11. <u>Other child development tests used:</u> N/A
Achenbach Child Behavior Checklist	N/A	Turkey	children 4-5 years old (older than 5 included in the study as well)	Validated Turkish version	"To determine the overall effect of multiple anesthetics on the psychology of children"	"The children in Group S underwent a total of 251 (11 +/- 7) GAs over 4-60 months. The incidence of psychopathology was nine and 10 children in groups S and C, respectively. The CBCL and CDI scores were parallel with a psychiatric diagnosis. Marital conflict scores were higher in Group S. Both chronic disease states affect psychology of children. Repeated anesthesia in addition to chronic disease does not seem to disturb the child's psychological health further when tentative and precautious approach modalities are undertaken."	Kayaalp L, Bozkurt P, Odabasi G, Dogangun B, Cavusoglu P, Bolat N, Bakan M.	2006	Kayaalp L, Bozkurt P, Odabasi G, Dogangun B, Cavusoglu P, Bolat N, Bakan M. Psychological effects of repeated general anesthesia in children. Paediatr Anaesth. 2006 Aug;16(8):822-7. <u>Other child development tests used:</u> DSM-IV criteria and Child Depression Inventory
Achenbach Child Behavior Checklist	N/A	Turkey, The Netherlands	4 to 18 years	Turkish, Dutch	Study comparing problem behaviors in 2,081 Dutch children, 3,127 Turkish children in Ankara and 833 Turkish immigrant children living in The Netherlands, aged 4-18 years.	Immigrant children scored higher than Ankara children on five CBCL scales. However, these differences were much smaller than those found between immigrant and Dutch children. Furthermore, immigrant children's Total Problem scores did not differ from those for Ankara children. The higher scores for Turkish children on the Anxious/Depressed scale compared with their Dutch peers may be explained by cultural differences in parental perception of children's problem behaviors, as well as the threshold for reporting them, or by cultural differences in the prevalence of problems, for instance as the result of cross-cultural differences in child-rearing practice.	Bengi-Arslan et al.	1997	Bengi-Arslan L, Verhulst FC, van der Ende J, Erol N. Understanding childhood (problem) behaviors from a cultural perspective: comparison of problem behaviors and competencies in Turkish immigrant, Turkish and Dutch children. Soc Psychiatry Psychiatr Epidemiol. Nov 1997;32(8):477-484.
Ages and Stages Questionnaire (ASQ)	N/A	Ecuador	children 3-61 months	Not specified in abstract	"To identify and describe the sociodemographic and nutritional characteristics associated with neurobehavioral development among young children living in three communities in the northeastern Andean region of Cayambe-Tabacundo, Ecuador."	"High frequencies of developmental delay were observed. Children 3 to 23 months old displayed delay in gross motor skills (30.1%), and children 48 to 61 months old displayed delay in problem-solving skills (73.4%) and fine motor skills (28.1%). A high frequency of both anemia (60.4%) and stunting (53.4%) was observed for all age groups. Maternal educational level was positively associated with communication and problem-solving skills, and monthly household income was positively associated with communication, gross motor, and problem-solving skills. The results suggest a high prevalence of developmental delay and poor child health in this population. Child health status and the child's environment may contribute to developmental delay in this region of	Handal AJ, Lozoff B, Breilh J, Harlow SD.	2007	Handal AJ, Lozoff B, Breilh J, Harlow SD. Sociodemographic and nutritional correlates of neurobehavioral development: a study of young children in a rural region of Ecuador. Rev Panam Salud Publica. 2007 May;21(5):292-300. <u>Other child development tests used:</u> N/A

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						Ecuador, but sociodemographic factors affecting opportunities for stimulation may also play a role. Research is needed to identify what is causing high percentages of neurobehavioral developmental delay in this region of Ecuador."			
Ages and Stages Questionnaire (ASQ)	N/A	Ecuador	children 3-61 months	Spanish version adapted to local vernacular. Culturally inappropriate language was removed and the Quichua word for child/baby was added.	"In this study we compared neurobehavioral development in Ecuadoran children living in two communities with high potential for exposure to organophosphate (OP) and carbamate pesticides to that of children living in a community with low potential for exposure."	"Children 3–23 months of age who resided in high-exposure communities scored lower on gross motor (p = 0.002), fine motor (p = 0.06), and socioindividual (p-value = 0.02) skills, compared with children in the low-exposure community. The effect of residence in a high-exposure community on gross motor skill development was greater for stunted children compared with non-stunted children (p = < 0.001) in the same age group of 3–23 months. Children 24–61 months of age residing in the high-exposure communities scored significantly lower on gross motor skills compared with children of similar ages residing in the low-exposure community (p = 0.06). Residence in communities with high potential for exposure to OP and carbamate pesticides was associated with poorer neurobehavioral development of the child even after controlling for major determinants of delayed development. Malnourished populations may be particularly vulnerable to neurobehavioral effects of pesticide exposure."	Handal AJ, Lozoff B, Breilh J, Harlow SD.	2007	Handal AJ, Lozoff B, Breilh J, Harlow SD. Effect of community of residence on neurobehavioral development in infants and young children in a flower-growing region of Ecuador. Environ Health Perspect. 2007 Jan;115(1):128-33. <u>Other child development tests used:</u> N/A
Ages and Stages Questionnaire (ASQ)	N/A	Ecuador	infants 3-23 months	Spanish version adapted to local vernacular. Culturally inappropriate language was removed.	to study "the potential effects of maternal occupation in the cut-flower industry during pregnancy on neurobehavioral development in Ecuadorian children."	"Children whose mothers worked in the flower industry during pregnancy scored lower on communication (8% decrease in score, 95% confidence interval [CI]: -16% to 0.5%) and fine motor skills (13% decrease, 95% CI: -22% to -5), and had a higher odds of having poor visual acuity (odds ratio = 4.7 [CI =1.1–20]), compared with children whose mothers did not work in the flower industry during pregnancy, after adjusting for potential confounders. Maternal occupation in the cut-flower industry during pregnancy may be associated with delayed neurobehavioral development of children aged 3-23 months. Possible hazards associated with working in the flower industry during pregnancy include pesticide exposure, exhaustion, and job stress."	Handal AJ, Harlow SD, Breilh J, Lozoff B.	2008	Handal AJ, Harlow SD, Breilh J, Lozoff B. Occupational exposure to pesticides during pregnancy and neurobehavioral development of infants and toddlers. Epidemiology. 2008 Nov;19(6):851-9. <u>Other child development tests used:</u> N/A
Ages and Stages Questionnaire (ASQ)	N/A	Ecuador	children 24-61 months	Spanish version adapted to local vernacular. Culturally inappropriate	"This preliminary study conducted in Ecuador examines the association between household and environmental risk factors for pesticide exposure and neurobehavioral	"Current maternal employment in the flower industry was associated with better developmental scores. Longer hours playing outdoors were associated with lower gross and fine motor and problem solving skills. Children who played with irrigation water scored lower on fine motor skills (8% decrease; 95% confidence interval = -9.31 to -	Handal AJ, Lozoff B, Breilh J, Harlow SD.	2007	Handal AJ, Lozoff B, Breilh J, Harlow SD. Neurobehavioral development in children with potential exposure to pesticides. Epidemiology. 2007 May;18(3):312-20.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
				language was removed and the Quichua word for child/baby was added.	development."	0.53), problem-solving skills (7% decrease; -8.40 to -0.39), and Visual Motor Integration test scores (3% decrease; -12.00 to 1.08). These results suggest that certain environmental risk factors for exposure to pesticides may affect child development, with contact with irrigation water of particular concern. However, the relationships between these risk factors and social characteristics are complex, as corporate agriculture may increase risk through pesticide exposure and environmental contamination, while indirectly promoting healthy development by providing health care, relatively higher salaries, and daycare options."			<u>Other child development tests used:</u> Visual Motor Integration Test
Australian Early Developmental Index	N/A	Australia	4 to 5 years	English	Study to examine the construct and concurrent validity of the AEDI.	Construct validity was moderate to high (depending on construct). Findings of concurrent validity are inconclusive since there is no criterion measure with which to assess the AEDI.	Brinkman SA	2007	Brinkman SA, Silburn S, Lawrence D, Goldfield S, Sayers M, & Oberklaid F. (2007). Investigating the validity of the Australian Early Development Index. <i>Early Education and Development</i> , 18(3), 427-451.
Bayley Scales of Infant Development (BSID)	II	Bosnia	Used between 6 and 24 months and then a three month follow up.	Not specified	The effect of iron therapy on mental and motor development in children suffering from iron deficiency anaemia.	Indexes of mental development before and 3 months after iron therapy for group of patients with severe and mild form of anemia (Hb < 95) were not significantly different (p > 0.05) before and after three months after iron therapy. There were statistically significant differences between 3 groups (Hb < 95 Hb = 95 -110 Hb; non-anemic) before therapy, however.	Hasanbegovic et al.	2004	Hasanbegovic E, Sabanovic S. [Effects of iron therapy on motor and mental development of infants and small children suffering from iron deficiency anaemia] <i>Med Arh</i> . 2004;58(4):227-9. Bosnian.
Bayley Scales of Infant Development (BSID)	II	Argentina	Children under 6 years	Spanish	A national psychomotor development survey was compared against the Bayley and other standardized measures.	Comparative test showed no significant differences between tests. Multiple logistic regressions showed that social class, maternal education and sex (female) were associated with earlier attainment of some selected developmental items, achieved at ages later than 1 year. Selected items achieved before the first year of life were not affected by any of the independent environmental variables studied.	Lejarraga et al.	2002	Lejarraga H, Pascucci MC, Krupitzky S, Kelmansky D, Bianco A, Martinez E, Tibaldi F, Cameron N. Psychomotor development in Argentinean children aged 0-5 years. <i>Paediatr Perinat Epidemiol</i> . 2002 Jan;16(1):47-60.
Bayley Scales of Infant Development (BSID)	II	Bangladesh	6 and 12 months	Not specified	Study to examine whether a weekly supplement of iron, zinc, iron+zinc, or a micronutrient mix (MM) of 16 vitamins and minerals would alter infant development and behavior.	When administered together, weekly iron and zinc supplementation improve motor development and orientation-engagement.	Black et al.	2004	Black MM, Baqui AH, Zaman K, Ake Persson L, El Arifeen S, Le K, McNary SW, Parveen M, Hamadani JD, Black RE. Iron and zinc supplementation promote motor development and exploratory behavior among Bangladeshi infants. <i>Am J Clin Nutr</i> . 2004 Oct;80(4):903-10.
Bayley Scales	II	Bangladesh	7 and 13	Not specified	Study to assess the effect of	The mental development index scores of the zinc-	Hamadani et al.	2001	Hamadani JD, Fuchs GJ,

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
of Infant Development (BSID)			months		zinc supplementation on the developmental levels and behavior of Bangladeshi infants.	treated group were slightly but significantly lower than those of the placebo group. This finding may have been due to micronutrient imbalance. Caution should be exercised when supplementing undernourished infants with a single micronutrient.			Osendarp SJ, Khatun F, Huda SN, Grantham-McGregor SM. Randomized controlled trial of the effect of zinc supplementation on the mental development of Bangladeshi infants. Am J Clin Nutr. 2001 Sep;74(3):381-6.
Bayley Scales of Infant Development (BSID)	II	Bangladesh	6-24 months	Not specified in abstract or methods.	"The aim of the study was to incorporate stimulation into the routine treatment of severely malnourished children in a nutrition unit and evaluate the impact on their growth and development."	"Twenty-seven children were lost to the study. In the remaining children, both groups had similar developmental scores and anthropometry initially. After 6 months, the intervention group had improved more than the controls did by a mean of 6.9 (P<0.001; 95% CI: 3.9, 10.0) mental and 3.1 (P=0.024; 95% CI: 0.4, 5.7) motor raw scores and a mean of 0.4 (P=0.029; 95% CI: 0.1, 0.8) weight-for-age z scores, controlling for background variables. Psychosocial stimulation integrated into treatment of severely malnourished children in hospital, followed by home visits for 6 months, was effective in improving children's growth and development and should be an integral part of their treatment."	Nahar B, Hamadani JD, Ahmed T, Tofail F, Rahman A, Huda SN, Grantham-McGregor SM.	2008	Nahar B, Hamadani JD, Ahmed T, Tofail F, Rahman A, Huda SN, Grantham-McGregor SM. Effects of psychosocial stimulation on growth and development of severely malnourished children in a nutrition unit in Bangladesh. Eur J Clin Nutr. 2008 Sep 3. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Bangladesh	infants at 6 and 12 months	Not specified in abstract or methods, used US norms	"To examine how maternal depressive symptoms are related to infant development among low-income infants in rural Bangladesh and to examine how the relationship is affected by maternal perceptions of infant irritability and observations of caregiving practices."	"Half (52%) the mothers reported depressive symptoms. In bivariate analyses, maternal depressive symptoms were related to low scores on the Bayley Scales. Infants whose mothers reported depressive symptoms and perceived their infants to be irritable acquired fewer cognitive, motor, and Orientation/Engagement skills between 6-12 months than infants whose mothers reported neither or only one condition. The relationship linking maternal depressive symptoms and perceived infant irritability with infant cognitive skills was partially mediated by parental responsiveness and opportunities for play in the home. The intergenerational risks of maternal depressive symptoms on infant development extend to rural Bangladesh and are accentuated when mothers perceive their infants as irritable. Mothers who report depressive symptoms and infant irritability may lack the capacity to provide responsive, developmentally-oriented caregiving environments."	Black MM, Baqui AH, Zaman K, McNary SW, Le K, Arifeen SE, Hamadani JD, Parveen M, Yunus M, Black RE.	2007	Black MM, Baqui AH, Zaman K, McNary SW, Le K, Arifeen SE, Hamadani JD, Parveen M, Yunus M, Black RE. Depressive symptoms among rural Bangladeshi mothers: implications for infant development. J Child Psychol Psychiatry. 2007 Aug;48(8):764-72. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Bangladesh	children 6-24 months	Not specified	"We added psychosocial stimulation to the treatment of undernourished children in a randomized controlled	"Intervention benefited children's mental development (4.6 +/- 2.0, P = 0.02), vocalization (0.48 +/- 0.23, P = 0.04), cooperation (0.45 +/- 0.16, P = 0.005), response-to-examiner (0.50 +/- 0.15, P =	Hamadani JD, Huda SN, Khatun F, Grantham-	2006	Hamadani JD, Huda SN, Khatun F, Grantham-McGregor SM. Psychosocial stimulation improves the development of

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					trial to assess the effects on children's development and growth and mothers' knowledge."	0.001), emotional tone (0.33 +/- 0.15, P = 0.03), and mothers' knowledge (3.5 +/- 0.49, P < 0.001). At the end, undernourished controls had poorer mental (-4.6 +/- 2.0, P = 0.02) and motor (-6.6 +/- 2.2, P = 0.003) development, were more inhibited (-0.35 +/- 0.16, P = 0.03), fussier (-0.57 +/- 0.16, P < 0.001), less cooperative (-0.48 +/- 0.17, P = 0.005), and less vocal (-0.76 +/- 0.23, P = 0.001) than better-nourished children. Intervened children scored lower only in motor development (-4.4 +/- 2.3, P = 0.049). Neither group of undernourished children improved in nutritional status, indicating that treatment had no effect. In conclusion, adding child development activities to the BINP improved children's development and behavior and their mothers' knowledge; however, the lack of improvement in growth needs to be examined further."	McGregor SM.		undernourished children in rural Bangladesh. J Nutr. 2006 Oct;136(10):2645-52. <u>Other child development tests used:</u> Behavior ratings used based on modifications of Wolke's scales and used previously in Bangladesh
Bayley Scales of Infant Development (BSID)	II	Brazil	17 to 42 months	Not specified	Study to assess the association between quality of stimulation in the family environment and child's cognitive development considering the impact of mother's schooling on the quality of stimulation.	There was a positive (beta=0.66) and statistically significant association between quality of stimulation in the family environment and child's cognitive development. Part of the effect was mediated by the mother's working circumstances and educational level.	Andrade SA et al.	2005	Andrade SA, Santos DN, Bastos AC, Pedromonico MR, de Almeida-Filho N, Barreto ML. Family environment and child's cognitive development: an epidemiological approach. Rev Saude Publica. 2005 Aug;39(4):606-11. Epub 2005 Aug 16. Portuguese
Bayley Scales of Infant Development (BSID)	II	Brazil	6 and 12 months	Not specified	Study to test whether zinc supplementation reduces the deficits in mental development and behavior that are found in term infants of low birth weight in the study population.	Zinc supplementation (5 mg/d) for eight weeks may reverse some of the poor behaviors, particularly responsiveness, exhibited by low birth weight infants. No amelioration of their mental and psychomotor deficits was found.	Ashworth et al.	1998	Ashworth A, Morris SS, Lira PI, Grantham-McGregor SM. Zinc supplementation, mental development and behavior in low birth weight term infants in northeast Brazil. Eur J Clin Nutr. 1998 Mar;52(3):223-7.
Bayley Scales of Infant Development (BSID)	II	Brazil	2 months	Not specified	Study to compare the neurodevelopment of full-term adequate (AGA) or small-for-gestational age (SGA) infants in the second month of life.	The SGA group Index Score (IS) was significantly lower in Mental and Motor Scales.	Goto et al. (24)	2005	Goto MM, Goncalves VM, Netto AA, Morcillo AM, Moura-Ribeiro MV. Neurodevelopment of full-term small-for-gestational age infants in the second month of life. Arq Neuropsiquiatr. 2005 Mar;63(1):75-82. Epub 2005 Apr 13. Portuguese.
Bayley Scales of Infant Development (BSID)	II	Brazil	6 and 12 months	Not specified	Study to (1) compare the mental and psychomotor development of low birth weight term (LBW-T) infants with that of	LBW-T infants had poorer development than ABW infants and differed in their behavior. There was an interaction between birth weight and the environment. LBW-T infants, but not ABW infants, were affected by the quality of stimulation in the	Grantham-McGregor et al.	1998	Grantham-McGregor SM, Lira PI, Ashworth A, Morris SS, Assuncao AM. The development of low birth weight term infants and the

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					appropriate birth weight (ABW) infants at 6 and 12 months of age; and to (2) examine the relationship between developmental levels and social background.	home and maternal illiteracy.			effects of the environment in northeast Brazil. J Pediatr. 1998 Apr;132(4):661-6.
Bayley Scales of Infant Development (BSID)	II	Brazil	6 and 12 months	Not specified	Study of the effects of morbidity and breastfeeding on the development of LBW term infants.	The infants' development was assessed on the Bayley Scales at 6 and 12 mo. Low birthweight group had lower scores than the higher birthweight group. Low birthweight infants are especially vulnerable to the effects of diarrhea, and the greater frequency and differential effect of diarrhea partly explains their poorer development.	Morris et al.	1999	Morris SS, Grantham-McGregor SM, Lira PI, Assuncao AM, Ashworth A. Effect of breastfeeding and morbidity on the development of low birthweight term babies in Brazil. Acta Paediatr. 1999 Oct;88(10):1101-6.
Bayley Scales of Infant Development (BSID)	adaptation	Brazil	Infants; age not quantified in abstract	Not specified in abstract. Visual behavior assessment of infants based on Bayley Scales of Infant Development	"We carried out a cross-sectional follow-up study of preterm infants in the first month of life who had their visuomotor behaviour evaluated at the chronological and corrected age."	"Most of the preterm infants presented a response, with a higher frequency in the eye contact tests, smiling as a social response, horizontal and vertical visual tracking, and increased mobility of the upper limbs on seeing the object at the corrected age. The responses obtained in this study allow us to confirm the importance of taking into account the corrected age when measuring the parameters involved in the development of visuomotor behaviour."	Albuquerque RC, Gagliardo HG, Lima AC, Guerra MQ, Rabelo AR, Cabral-Filho JE.	2009	Albuquerque RC, Gagliardo HG, Lima AC, Guerra MQ, Rabelo AR, Cabral-Filho JE. [Visuomotor behaviour of preterm infants in the first month of life. A comparison between the chronological and corrected ages]. Rev Neurol. 2009 Jan 1-15;48(1):13-6. <u>Other child development tests used:</u> Study used a visual behavior assessment based on Bayley Scales of Infant Development
Bayley Scales of Infant Development (BSID)	II	Brazil	infants at 6 months	Not specified in abstract or methods	"To ascertain the degree of agreement between a score for screening and another for diagnosis of motor development in 6-month old infants and to define the most appropriate cutoff point for screening."	"The study sample comprised 43 infants. Six infants (14.00%) exhibited inadequate motor performance. Using the BSID-II motor classification and the 5th percentile AIMS cutoff, sensitivity was 100%, specificity 78.37%, accuracy 81.39%, kappa index 0.50 and p < 0.001; whereas, using the BSID-II motor classification and the 10th percentile AIMS cutoff, sensitivity was 100%, specificity 48.64%, accuracy 55.81%, kappa index 0.20 and p 0.025. The results suggest that concordance between the two 6-month assessment scales is good. The parameters employed are best combined using the 5th percentile AIMS cutoff point."	Campos D, Santos DC, Gonçalves VM, Goto MM, Arias AV, Brianeze AC, Campos TM, Mello BB.	2006	Campos D, Santos DC, Gonçalves VM, Goto MM, Arias AV, Brianeze AC, Campos TM, Mello BB. Agreement between scales for screening and diagnosis of motor development at 6 months. J Pediatr (Rio J). 2006 Nov-Dec;82(6):470-4. <u>Other child development tests used:</u> Alberta Infant Motor Scale (AIMS)
Bayley Scales of Infant Development (BSID)	II	Brazil	infants at 1, 2, 3, and 6 months	Not specified in abstract	"To compare the motor performance of infants born small for gestational age (SGA) with those	"The SGA group presented a mean motor index score lower than the AGA group at 2 and 6 months, with the SGA group presenting fewer infants that successfully accomplished "makes crawling	Campos D, Santos DC, Gonçalves VM, Goto MM,	2008	Campos D, Santos DC, Gonçalves VM, Goto MM, Campos-Zanelli TM. Motor performance of infants born

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					appropriate for gestational age (AGA) at 1, 2, 3, and 6 months."	movements," "turns from side to back," "balances head," "sits alone momentarily," and "sits alone for 30 seconds. Data analysis suggested that infants who are SGA present greater risk of adverse outcomes that are detectable in motor performance measures at 2 months."	Campos-Zanelli TM		small or appropriate for gestational age: a comparative study. <i>Pediatr Phys Ther.</i> 2008 Winter;20(4):340-6. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Brazil	infants assessed at 6 or 12 months	Not specified in abstract or methods. Note: article states that BSID had not undergone transcultural adaptation for Brazil at the time of publication but nevertheless is widely used	"To verify the concurrent validity and interobserver reliability of the Alberta Infant Motor Scale (AIMS) in premature infants followed-up at the outpatient clinic of Instituto Fernandes Figueira, Fundação Oswaldo Cruz (IFF/Fiocruz), in Rio de Janeiro, Brazil."	"The concurrent validity study found a high level of correlation between the two scales ($r = 0.95$) and one that was statistically significant ($p < 0.01$) for the entire population of infants, with higher values at 12 months ($r = 0.89$) than at 6 months ($r = 0.74$). The interobserver reliability study found satisfactory intraclass correlation coefficients at all ages tested, varying from 0.76 to 0.99. The AIMS is a valid and reliable instrument for the evaluation of motor development in high-risk infants within the Brazilian public health system."	Almeida KM, Dutra MV, de Mello RR, Reis AB, Martins PS.	2008	Almeida KM, Dutra MV, de Mello RR, Reis AB, Martins PS. Concurrent validity and reliability of the Alberta Infant Motor Scale in premature infants. <i>J Pediatr (Rio J).</i> 2008 Sep-Oct;84(5):442-8. <u>Other child development tests used:</u> AIMS
Bayley Scales of Infant Development (BSID)	II	Brazil	children 20-42 months	Not specified in abstract or methods	"The study describes the relationship between anthropometric status, socioeconomic conditions, and quality of home environment and child cognitive development in 320 children from 20 to 42 months of age, randomly selected from 20,000 households that represent the range of socioeconomic and environmental conditions in Salvador, Bahia, Northeast Brazil."	"Socioeconomic factors were found to have an indirect impact on early cognitive development mediated by the child's proximal environment factors, such as appropriate play materials and games available and school attendance. No independent association was seen between nutritional status and early cognitive development."	Marques dos Santos L, Neves dos Santos D, Bastos AC, Assis AM, Prado MS, Barreto ML.	2008	Marques dos Santos L, Neves dos Santos D, Bastos AC, Assis AM, Prado MS, Barreto ML. Determinants of early cognitive development: hierarchical analysis of a longitudinal study. <i>Cad Saude Publica.</i> 2008 Feb;24(2):427-37. <u>Other child development tests used:</u> Home Observation for Measurement of the Environmental Inventory (HOME)
Bayley Scales of Infant Development (BSID)	II	Brazil	children at 12 and 18 months	Not specified in abstract or methods	"To investigate growth and neurodevelopment outcome of very low birth weight (VLBW) infants delivered by preeclamptic mothers."	"40 infants in preeclamptic and 46 in control groups were studied. Birth weight and gestational age were 1148 g+/-236 and 1195 g+/-240, and 31.3 weeks+/-1 and 30.6 weeks+/-2 for preeclamptic and control groups, respectively. At 12 and 18 months, CA, weight for age (Z score) was significantly higher in control than in preeclamptic. PDI scores were higher in preeclamptic than in controls at 18 months CA. Catch-up of body weight did not occur in the first 18 months CA in preeclamptic infants. Neurodevelopment outcome was better in infants delivered by preeclamptic mothers than in controls	Silveira RC, Procianoy RS, Koch MS, Benjamin AC, Schlindwein CF.	2007	Silveira RC, Procianoy RS, Koch MS, Benjamin AC, Schlindwein CF. Growth and neurodevelopment outcome of very low birth weight infants delivered by preeclamptic mothers. <i>Acta Paediatr.</i> 2007 Dec;96(12):1738-42. Epub 2007 Oct 22. <u>Other child development tests used:</u>

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						at 18 months CA."			N/A
Bayley Scales of Infant Development (BSID)	II	Brazil	children 12 months	Not specified in abstract or methods	"The aim of this prospective study was to investigate the association between breast feeding and mental and motor development at age 12 months, controlling for comprehensive measures of the child's socio-economic, maternal and environmental background, and nutritional status. "	"After adjusting for potential confounders, full breast feeding (i.e. exclusive or predominant) at 1 month was associated with a small but significant benefit in mental development (+3.0 points, P = 0.02) compared with partial or no breast feeding. No additional advantage in mental development was found with longer durations of full breast feeding. Full breast feeding at 1 month was associated with improved behaviour for two of 10 ratings tested: initiative with tasks (P = 0.003) and attention (P = 0.02). No association between breast feeding and motor development was found."	Eickmann SH, de Lira PI, Lima Mde C, Coutinho SB, Teixeira Mde L, Ashworth A.	2007	Eickmann SH, de Lira PI, Lima Mde C, Coutinho SB, Teixeira Mde L, Ashworth A. Breast feeding and mental and motor development at 12 months in a low-income population in northeast Brazil. Paediatr Perinat Epidemiol. 2007 Mar;21(2):129-37. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Chile	12 months	Spanish	Study to evaluate relationship between breast feeding (BF) and psychomotor development (PMD) in two groups with different BF duration (1) 6 months; and (2) before 45 days.	Mean Mental Development Index (MDI) and Psychomotor Developmental Index (PDI) were similar in both groups.	de Andraca et al.	1999	de Andraca I, Salas MI, Lopez C, Cayazzo MS, Icaza G. Effect of breast feeding and psychosocial variables upon psychomotor development of 12-month-old infants. Arch Latinoam Nutr. 1999 Sep;49(3):223-31. Spanish.
Bayley Scales of Infant Development (BSID)	II	Chile	Between 6 and 12 months	Spanish	Study to evaluate the effect of risk factors on infant development, among low socioeconomic children born under optimal biological conditions, and who are exposed to adverse social circumstances.	The children's mental (MDI) and psychomotor (PDI) development was evaluated using the Bayley Scale of Infant Development. Eighteen risk factors were identified and dichotomized (high or low risk). The findings suggest that even for children born under optimal biological conditions their psychomotor development is negatively affected by the presence of simultaneous adverse environmental conditions.	de Andraca et al.	1998	de Andraca I, Pino P, de la Parra A, Rivera F, Castillo M. Risk factors for psychomotor development among infants born under optimal biological conditions. Rev Saude Publica. Apr 1998;32(2):138-47.
Bayley Scales of Infant Development (BSID)	II	Chile	At 6 months and every 6 months to 24 months	Spanish	Study to determine the effects of lead exposure, comparing Chilean lactating children residing in rural areas with low lead exposure, and in urban areas with high lead exposure.	In this cohort of children, no effect of lead levels on neurobehavioral development was found.	Vega et al.	1999	Vega J, Frenz P, Marchetti N, Torres J, Kopplin E, Delgado I, Vega F. Chronic exposure to environmental lead in Chilean infants. II: Effects on the psychomotor development] Rev Med Chil. 1999 Jan;127(1):28-37. Spanish.
Bayley Scales of Infant Development (BSID)	II	China	Younger than 2 1/2 years.	Chinese	Study to examine whether BSID is a good assessment for intelligence among at risk children younger than 2 1/2 years.	The development indices of BSID has a better predictive power on the intelligence among at-risk group u-Cheng children than the Stanford-Binet Intelligence Quotients (S-B IQ).	Chen et al.	1993	Chen YC, Hu HF, Lo HY, Hsu CC. [Correlation between development indices and later intelligence quotients-the follow-up study of chlorinated biphenyls affected children] Zhonghua Min Guo Xiao Er Ke Yi Xue Hui Za Zhi. 1993 Jul-Aug;34(4):278-84.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Bayley Scales of Infant Development (BSID)	II	China	3, 6, 12 months and 6 years	Chinese	Study to examine prognosis of SEC infants after 6 years (physical development, mental development, brain-stem auditory evoked potential (BAEP) and vision).	The parameters of the physical development in SEC infants (those with subependymal cysts) reached the similar level as controls till 6 years old. However, the index of mental development below 80 was more often seen in infants with SEC comparing to that in control group during the six year study period.	Qian et al.	2004	Qian JH, Chen HJ, Chen GY, Chen XP, Zhang Q, Ao LM, Wu SM. Prospective study on prognosis of infants with neonatal subependymal cysts. Zhonghua Er Ke Za Zhi. 2004 Dec;42(12):913-6. Chinese
Bayley Scales of Infant Development (BSID)	II	Costa Rica	Between 2 and 20 months	Spanish	Cross-sectional study to assess the nutritional status, cognitive development, and mother-child interactions in a group of 153 Nicaraguan refugee children living in Costa Rica.	There was no significant correlation between the total amount of mother-child interaction and child cognitive development. However, certain aspects of the home environment correlated with cognitive development, specifically the manner in which the mother responded emotionally and verbally to her child, and the organization of the child's physical and temporal environment. The manner in which the mother responded and the child's weight-for-height also were important in predicting child cognitive development. The child's weight-for-height and certain aspects of the home environment also played an important role in the cognitive development.	Laude et al.	1999	Laude M. Assessment of nutritional status, cognitive development, and mother-child interaction in Central American refugee children. Rev Panam Salud Publica. 1999 Sep;6(3):164-71.
Bayley Scales of Infant Development (BSID)	II	Costa Rica	1 week and 3 months	Spanish	Study to examine association between iron-deficiency anemia (IDA) and infant behavior and development.	Infants with IDA showed significantly lower mental and motor test scores, even after factors relating to birth, nutrition, family background, parental IQ, and the home environment were considered.	Lozoff et al.	1989	Lozoff B. Methodologic issues in studying behavioral effects of infant iron-deficiency anemia. Am J Clin Nutr. 1989 Sep;50(3 Suppl):641-51; discussion 652-4.
Bayley Scales of Infant Development (BSID)	II	Czech Republic	11 and 15 months	Czech	Study to evaluate influence of early nutrition on growth parameters and psychomotor development of children with very low birth weight.	Breast milk fortification had a favorable (statistically significant) effect on VLBW newborns, especially on motor development.	Kocourkova et al.	2004	Kocourkova I, Sobotkova D, Pilarova M, Dittrichova J, Vondracek J, Stranak Z. Effect of early nutrition on growth parameters and psychomotor development of children of very low birth weight. Ceska Gynekol. 2004 Dec;69 Suppl 1:108-13. Czech.
Bayley Scales of Infant Development (BSID)	II	Democratic Republic of the Congo	18-72 months	Lingala	"The purpose of this study was to compare the neurodevelopment of preschool-aged HIV-infected, HIV-affected (HIV-uninfected AIDS orphans and HIV-uninfected children whose mother had symptomatic AIDS), and healthy control children in Kinshasa, Democratic Republic of Congo."	"Overall, 60% of HIV-infected children had severe delay in cognitive function, 29% had severe delay in motor skills, 85% had delays in language expression, and 77% had delays in language comprehension, all significantly higher rates as compared with control children. Young HIV-infected children (aged 18-29 months) performed worse, with 91% and 82% demonstrating severe mental and motor delay, respectively, compared with 46% and 4% in older HIV-infected children (aged 30-72 months). HIV-affected children had significantly more motor and language expression	Van Rie A, Mupuala A, Dow A.	2008	Van Rie A, Mupuala A, Dow A. Impact of the HIV/AIDS epidemic on the neurodevelopment of preschool-aged children in Kinshasa, Democratic Republic of the Congo. Pediatrics. 2008 Jul;122(1):e123-8. <u>Other child development tests used:</u> Peabody Developmental Motor

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						delay than control children. The impact of the HIV pandemic on children's neurodevelopment extends beyond the direct effect of the HIV virus on the central nervous system. AIDS orphans and HIV-negative children whose mothers had AIDS demonstrated significant delays in their neurodevelopment, although to a lesser degree and in fewer developmental domains than HIV-infected children. Young HIV-infected children were the most severely afflicted group, indicating the need for early interventions. Older children performed better as a result of a "survival effect," with only those children with less aggressive disease surviving."			Scales, Snijders-Oomen Nonverbal Intelligence Test, Rossetti Infant-Toddler Language Scale
Bayley Scales of Infant Development (BSID)	II	Egypt	6 months	Not specified	Study to examine association between zinc intake of mothers during the last 6 months of pregnancy and the first 6 months of lactation, and child development outcomes.	Performance on the Bayley motor test at 6 mo of age was negatively related to maternal intakes of plant zinc, phytate, and fiber, suggesting that zinc bioavailability was involved.	Kirksey et al.G19	1994	Kirksey A, Wachs TD, Yunis F, Srinath U, Rahmanifar A, McCabe GP, Galal OM, Harrison GG, Jerome NW. Relation of maternal zinc nutriture to pregnancy outcome and infant development in an Egyptian village. Am J Clin Nutr. 1994 Nov;60(5):782-92.
Bayley Scales of Infant Development (BSID)	II	Ethiopia	16 to 42 months	Amharic	A study to determine the mental development of the children using the Bayley Scale of Mental Development, and to examine its relation to nutritional status and mother-child interaction.	The child's weight for age was significantly related to the child's scores on the Bayley scale. The rate of a mother's verbal responses to the child positively predicted the child's verbal score. In contrast, the mother's spontaneously initiated motor actions toward the child correlated negatively with the child's performance score.	About et al.	1995	About FE; Alemu T. Nutrition, Maternal Responsiveness and Mental Development of Ethiopian Children. Social Science and Medicine, Sep 1995; 41(5): 725-732.
Bayley Scales of Infant Development (BSID)	II	Ethiopia	2 years through adulthood	Amharic	Study to examine the possibility that there is an early sensitive period for the effects of malnutrition on cognitive development. Study included three groups (early growth falterers, late growth falterers, and controls).	Mean (SD) scores on the psychomotor scale were 10.2 (3.7) in the controls, 6.6 (4.2) in the early growth falterers, and 8.5 (4.3) in the late growth falterers. For the mental scale they were 28.9 (5.8), 22.6 (6.2), and 26.6 (6.1) respectively.	Drewett et al.	2001	Drewett R, Wolke D, Asefa M, Kaba M, Tessema F. Malnutrition and mental development: is there a sensitive period? A nested case-control study. J Child Psychol Psychiatry. 2001. Feb;42(2):181-7.
Bayley Scales of Infant Development (BSID)	II	Guatemala	Between 12 and 24 months	Spanish	Study to assess the affects of discontinuing coffee intake on iron deficient Guatemalan toddlers' cognitive development and sleep.	The effects of postnatal coffee ingestion in Guatemala were seen for sleep duration, but not for cognitive development. Prenatal coffee ingestion was negatively associated with Behavior Rating Scales and should be investigated further.	Engle et al.	1999	Engle PL, VasDias T, Howard I, Romero-Abal ME, Quan de Serrano J, Bulux J, Solomons NW, Dewey KG. Effects of discontinuing coffee intake on iron deficient Guatemalan toddlers' cognitive development and sleep. Early Hum Dev. Jan

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
									1999;53(3):251-69.
Bayley Scales of Infant Development (BSID)	II	Guatemala	children at 6, 24, and 36 months	Spanish language adaptation, pretested in population	"The objective of this study was to assess the relationship between growth and neurodevelopment during early childhood (birth-36 months)."	"Birth size was significantly associated with child development at 6 and 24 months. Gains in length and weight during the first 24 months were positively associated with child development, whereas growth from 24 to 36 months age was not associated with child development at 36 months. Motor development was more strongly and consistently related to child growth than was mental development. Head circumference gain after 6 months was not a significant predictor of child development at 24 and 36 months. Small size at birth and poor physical growth during the first 24 months are related to neurodevelopmental delays. More evidence from developing countries will help explain the underlying mechanisms and identify appropriate interventions to prevent neurodevelopmental delay in early childhood."	Kuklina EV, Ramakrishnan U, Stein AD, Barnhart HH, Martorell R.	2006	fKuklina EV, Ramakrishnan U, Stein AD, Barnhart HH, Martorell R. Early childhood growth and development in rural Guatemala. Early Hum Dev. 2006 Jul;82(7):425-33. Epub 2006 Jan 23. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Indonesia	12 and 18 months, tested at baseline, 2, 4, and 6 months	Indonesian?	Study about the effects of an energy and micronutrient supplementation on mental and motor development, activity and behavior under natural conditions of iron deficient anemic and iron replete children.	Anemic children showed faster motor development and greater physical activity than the control children did.	Harahap et al.	2000	Harahap H, Jahari AB, Husaini MA, Saco-Pollitt C, Pollitt E. Effects of an energy and micronutrient supplement on iron deficiency anemia, physical activity and motor and mental development in undernourished children in Indonesia. Eur J Clin Nutr. 2000 May;54 Suppl 2:S114-9.
Bayley Scales of Infant Development (BSID)	II	Indonesia	12-18-months	Not specified	A double-blind trial was done to monitor the effects of iron supplementation on performance in the Bayley scales of mental and motor development among 12-18-month-old infants.	After intervention, developmental delays were reversed among iron-deficient anaemic infants who had received iron but they remained the same among placebo-treated iron-deficient anaemic infants. Treatment with ferrous sulphate can improve the performance of 12-18-month-old iron-deficient anaemic infants on the Bayley scales of mental and motor development.	Idjradinata P & Politt E. et al.	1993	Idjradinata P, Pollitt E. Reversal of developmental delays in iron-deficient anaemic infants treated with iron. Lancet. 1993 Jan 2;341(8836):1-4.
Bayley Scales of Infant Development (BSID)	II	Indonesia	12 and 18 months, tested at baseline, 2, 4, and 6 months	Not specified	Study about the effects of an energy and micronutrient supplementation on mental and motor development, activity and behavior under natural conditions of iron deficient anemic and iron replete children.	Same study as 32. Children that received the E supplement walked at an earlier age, had higher scores in the Bayley Scale and were motorically more active. Similar intergroup differences were observed in the 18-month-old cohort in the total motor activity score.	Jahari et al.	2000	Jahari AB, Saco-Pollitt C, Husaini MA, Pollitt E. Effects of an energy and micronutrient supplement on motor development and motor activity in undernourished children in Indonesia. Eur J Clin Nutr. 2000 May;54 Suppl 2:S60-8.
Bayley Scales of Infant Development (BSID)	II	Indonesia	6 and 12 months	Not specified	Study to compare the effects of combined iron and zinc supplementation in infancy with the effects of iron and	Single supplementation with zinc significantly improved growth. Single supplementation with iron significantly improved growth and psychomotor development. Combined supplementation with iron	Lind et al.	2004	Lind T, Lonnerdal B, Stenlund H, Gamayanti IL, Ismail D, Seswandhana R, Persson LA. A community-based

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					zinc as single micronutrients on growth, psychomotor development, and incidence of infectious disease.	and zinc had no significant effect on growth or development. Combined iron and zinc supplementation to infants cannot be routinely recommended at the iron-to-zinc ratio used in this study.			randomized controlled trial of iron and zinc supplementation in Indonesian infants: effects on growth and development. <i>Am J Clin Nutr.</i> 2004 Sep;80(3):729-36.
Bayley Scales of Infant Development (BSID)	II	Indonesia	6 and 12 months	Bahasa Indonesia	A study to investigate the effects of vitamin A and Fe supplementation during gestation on infant mental and psychomotor development.	the present study did not find an impact of weekly supplementation of 4800 RE vitamin A in addition to Fe during gestation on functional development of Indonesian infants. However, smaller improvements in development may be seen if studied in a larger and/or more deficient population.	Schmidt MK et al.	2004	Schmidt MK, Muslimatun S, West CE, Schultink W, Hautvast JG. Mental and psychomotor development in Indonesian infants of mothers supplemented with vitamin A in addition to iron during pregnancy. <i>Br J Nutr.</i> 2004 Feb;91(2):279-86.
Bayley Scales of Infant Development (BSID)	II	Italy	One year	Italian	Study to examine association between duration of breastfeeding and developmental outcomes for children at age 1.	Children breastfed more than 6 months did better on the Bayley psychomotor developmental index than those breast feed more than 3 months but less than 6 months.	Agostini et al.	2001	Agostoni C, Marangoni F, Lammardo AM, Giovannini M, Riva E, Galli C. Breastfeeding duration, milk fat composition and developmental indices at 1 year of life among breastfed infants. <i>Prostaglandins Leukot Essent Fatty Acids.</i> 2001 Feb;64(2):105-9.
Bayley Scales of Infant Development (BSID)	II	Italy	3, 6, 9, 12, 24, 36 months	Italian	Study examined small for gestational age (SGA) born from 1980 to 1987 in Pavia. Bayley was used after 1985 to measure sychomotor development (Brunet-Lezine's Scale was used to measure until 1985).	At 36 months, SGA infants showed transient neurological abnormalities more frequently than the control group (30.7% vs 6.8% - $p < 0.001$). 5 SGA (9.6%) and 10 AGA (appropriate for gestational age) (13.7%) had minor abnormalities (group B); no SGA children and only one AGA had diplegia (group C); 3 SGA (5.8%) and 4 AGA (5.5%) were considered to have severe handicap (group D) SGA children had a higher incidence of epilepsy (3.8% vs 0) than AGA (group E).	Fazzi et al.	1992	Fazzi E, Orcesi S, Spinillo A, Stronati M, Telesca C, Farinotti L. . Neuropsychologic development of small for gestational age preterm infants: follow up at 12-36 months of age. <i>Pediatr Med Chir.</i> 1992 Jul-Aug;14(4):403-7. Italian.
Bayley Scales of Infant Development (BSID)	II	Jamaica	Birth to 5 years?	Not specified	Study to evaluate the effects of cannabis consumption during pregnancy and lactation on infants from birth to school age in rural communities.	Most scores fell in the middle range of about 4, similar to the North American scores, except for the lower mean in the category of Threshold of Responsiveness, because of an unanticipated cultural difference. The adjustments made did not compromise the comparability of the findings.	Dreher et al.	1993	Dreher MC, Hayes JS. Triangulation in cross-cultural research of child development in Jamaica. <i>West J Nurs Res.</i> 1993 Apr;15(2):216-29.
Bayley Scales of Infant Development (BSID)	II	Japan	7 & 18 months	Japanese	Study to evaluate associations between the neurobehavioral development of children and perinatal exposures to (Methylmercury) MeHg and environmentally organic	Results not reported yet; will follow children up to ages 6 & 7.	Nakai et al. (15)	2004	Nakai K, Suzuki K, Oka T, Murata K, Sakamoto M, Okamura K, Hosokawa T, Sakai T, Nakamura T, Saito Y, Kurokawa N, Kameo S, Satoh H. The Tohoku Study of Child Development: A cohort study

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					pollutants (POPs) in Japan.				of effects of perinatal exposures to methylmercury and environmentally persistent organic pollutants on neurobehavioral development in Japanese children. <i>Tohoku J Exp Med.</i> 2004 Mar;202(3):227-37.
Bayley Scales of Infant Development (BSID)	II	Kenya	6 months	Not specified	An empirical analysis of the factors affecting growth and psychological development of over 100 infants from birth to age 6 months in the Embu region of Kenya.	Infants' scores at 6 months on the Bayley Motor Scale and on eight items from the Bayley Infant Behavior Record were explained using anthropometric, socioeconomic, and psychological variables. The infants' arm circumference and intake of protein were significant predictors of scores on the Bayley Motor Scale. In addition, time spent by the mother talking to the infant was positively associated with the scores on the Bayley Infant Behavior Record.	Bhargava et al.	2000	Bhargava A. Modeling the effects of maternal nutritional status and socioeconomic variables on the anthropometric and psychological indicators of Kenyan infants from age 0-6 months. <i>Am J Phys Anthropol.</i> 2000 Jan;111(1):89-104.
Bayley Scales of Infant Development (BSID)	I or II (not indicated in abstract)	Lithuania	infants assessed at 1,3, 6, 9, and 12 months	Not specified in abstract	"The aim of this study was to assess psychomotor development of very-low-birth-weight infants."	"The results showed that at the age of 1 month, a significantly delayed psychomotor development was noted in 15.2% of preterm infants and in none of term infants; moderately delayed development - in 25.3% of preterm infants and in none of term infants; normal - in 54.4% and 90.3%, respectively; and accelerated development - 5.1% and 9.7%, respectively. At the age of 12 months, 21.8% of preterm infants and none of term infants showed a significantly delayed psychomotor development; 26.9% of study group patents and none in control group - moderately delayed development; 51.3% and 93.5%, respectively - normal development; and 6.5% and none, respectively - accelerated development. Psychomotor development of preterm infants is retarded during all first year of life."	Rimdeikiene I, Krisciūnas A, Markūniene E.	2008	Rimdeikiene I, Krisciūnas A, Markūniene E. [The evaluation of psychomotor development in preterm infants]. <i>Medicina (Kaunas).</i> 2008;44(5):378-85. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Malaysia	One year	Not specified	Study to determine neonatal, early developmental and social risk factors that predict the neurocognitive and behavioural outcome of very low birthweight (VLBW) preschool children at four years of age. Cognitive development was assessed using the Mental Scale of the Bayley Scales of Infant Development (MDI) at one year and the Weschler Preschool and Primary Scale of	Social factors and the caregiving environment were important determinants of cognitive and behavioural outcome. (Outcomes were measured with other scales; not the Bayley.)	Ong et al.	2001	Ong LC, Boo NY, Chandran V. Predictors of neurodevelopmental outcome of Malaysian very low birthweight children at 4 years of age. <i>J Paediatr Child Health.</i> 2001 Aug;37(4):363-8.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					Intelligence-Revised (WIPPSI-R) at four years.				
Bayley Scales of Infant Development (BSID)	II	Mexico	3 and 6 months	Spanish	Study to examine the relationship between nutritional conditions of the mother during pregnancy and the child and the psychomotor development of young infants from a poor and underfed rural community.	Small variations in nutrition of the mother-child pairs from the poor rural community studied had consequences on the psychomotor development of the infant.	Chavez et al.	1998	Chavez A, Martinez H, Guarneros N, Allen L, Pelto G. Nutrition and development in the first semester of life. Salud Publica Mex. 1998 Mar-Apr;40(2):111-8. Spanish
Bayley Scales of Infant Development (BSID)	II	Mexico	12.5-23.5 months	Spanish	Cross-sectional study anthropometric indicators and mental development in very-low-income children in the second year of life.	There was a significant decline in HAZ and in age-adjusted MDI score across the second year of life. Although the children showed MDI scores close to the mean, normed US values at 13-14 months, the scores were significantly lower than expected in older children (P < 0.0001), even after controlling for socio-economic status and parental characteristics. At 13-14 months, only 3% of children received scores below 70 (less than minus two standard deviations), whereas by 19-20 months, almost 17% of children were performing below this level.	Fernald et al.	2006	Fernald LC, Neufeld LM, Barton LR, Schnaas L, Rivera J, Gertler PJ. Parallel deficits in linear growth and mental development in low-income Mexican infants in the second year of life. Public Health Nutr. 2006 Apr;9(2):178-86.
Bayley Scales of Infant Development (BSID)	II	Mexico	2 to 54 weeks	Spanish	Study to assess motor development of infants in Yucatan as compared to U.S. infants.	In comparison to USA infants, early acceleration of motor development was followed by a marked downward trend. This phenomenon, if observed in a single child, may indicate progressive neurologic disease. Child-rearing practices would appear to account for the difference in pattern of test performance.	Solomons et al.	1978	Solomons HC. The malleability of infant motor development: cautions based on studies of child-rearing practices in Yucatan. Clin Pediatr (Phila). 1978 Nov;17(11):836-40.
Bayley Scales of Infant Development (BSID)	II	Mexico	infants 1-12, administered at 1, 3, 6, and 12 months	Spanish	"Our goal was to assess the prenatal DDE exposure window and its effect on the psychomotor development index (PDI) and mental development index (MDI) during the first year of life."	"Third-trimester DDE level (7.8 +/- 2.8 ppb) was significantly higher than the level at baseline, first, and second trimesters, but the differences never exceeded 20%. Only DDE levels during the first trimester of pregnancy were associated with a significant reduction in PDI (every doubled increase of DDE level reduced the PDI 0.5 points). DDE was not associated with MDI. A critical window of exposure to DDE in utero may be the first trimester of the pregnancy, and psychomotor development is a target of this compound. Residues of DDT metabolites may present a risk of developmental delay for years after termination of DDT use."	Torres-Sánchez L, Rothenberg SJ, Schnaas L, Cebrián ME, Osorio E, Del Carmen Hernández M, García-Hernández RM, Del Río-García C, Wolff MS, López-Carrillo L	2007	Torres-Sánchez L, Rothenberg SJ, Schnaas L, Cebrián ME, Osorio E, Del Carmen Hernández M, García-Hernández RM, Del Río-García C, Wolff MS, López-Carrillo L. In utero p,p'-DDE exposure and infant neurodevelopment: a perinatal cohort in Mexico. Environ Health Perspect. 2007 Mar;115(3):435-9. Epub 2007 Jan 16. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Mexico	children 12 and 24 months	Spanish	"We studied prenatal lead exposure's impact on	"Maternal lead levels were moderately high with a first-trimester blood lead mean (+/- SD) value of 7.1	Hu H, Téllez-Rojo MM,	2007	Hu H, Téllez-Rojo MM, Bellinger D, Smith D, Ettinger

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Development (BSID)					neurodevelopment using repeated measures of fetal dose as reflected by maternal whole blood and plasma lead levels. "	+/- 5.1 microg/dL and 14% of values ≥ 10 microg/dL. Both maternal plasma and whole blood lead during the first trimester (but not in the second or third trimester) were significant predictors ($p < 0.05$) of poorer Mental Development Index (MDI) scores. In models combining all three trimester measures and using standardized coefficients, the effect of first-trimester maternal plasma lead was somewhat greater than the effect of first-trimester maternal whole blood lead and substantially greater than the effects of second- or third-trimester plasma lead, and values averaged over all three trimesters. A 1-SD change in first-trimester plasma lead was associated with a reduction in MDI score of 3.5 points. Postnatal blood lead levels in the offspring were less strongly correlated with MDI scores. Fetal lead exposure has an adverse effect on neurodevelopment, with an effect that may be most pronounced during the first trimester and best captured by measuring lead in either maternal plasma or whole blood."	Bellinger D, Smith D, Ettinger AS, Lamadrid-Figueroa H, Schwartz J, Schnaas L, Mercado-García A, Hernández-Avila M.		AS, Lamadrid-Figueroa H, Schwartz J, Schnaas L, Mercado-García A, Hernández-Avila M. Fetal lead exposure at each stage of pregnancy as a predictor of infant mental development. Environ Health Perspect. 2006 Nov;114(11):1730-5. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Mexico	infants at 12 and 24 months	Instructions and prompts translated into Spanish	"Using data from a prospective study conducted in Mexico City, Mexico, we evaluated the dose-effect relationship between blood lead levels and neurodevelopment at 12 and 24 months of age."	"Adjusting for covariates, children's blood lead levels at 24 months were significantly associated, in an inverse direction, with both Mental Development Index and Psychomotor Development Index scores at 24 months. Blood lead level at 12 months of age was not associated with concurrent Mental Development Index or Psychomotor Development Index scores or with Mental Development Index at 24 months of age but was significantly associated with Psychomotor Development Index score at 24 months. The relationships were not altered by adjustment for cord blood lead level or, in the analyses of 24-month Mental Development Index and Psychomotor Development Index scores, for the 12-month Mental Development Index and Psychomotor Development Index scores. For both Mental Development Index and Psychomotor Development Index at 24 months of age, the coefficients that were associated with concurrent blood lead level were significantly larger among children with blood lead levels < 10 microg/dL than it was among children with levels > 10 microg/dL. These analyses indicate that children's neurodevelopment is inversely related to their blood lead levels even in the range of < 10 microg/dL. Our findings were consistent with a supralinear relationship between blood lead levels and neurobehavioral outcomes."	Téllez-Rojo MM, Bellinger DC, Arroyo-Quiroz C, Lamadrid-Figueroa H, Mercado-García A, Schnaas-Arrieta L, Wright RO, Hernández-Avila M, Hu H.	2006	Téllez-Rojo MM, Bellinger DC, Arroyo-Quiroz C, Lamadrid-Figueroa H, Mercado-García A, Schnaas-Arrieta L, Wright RO, Hernández-Avila M, Hu H. Longitudinal associations between blood lead concentrations lower than 10 microg/dL and neurobehavioral development in environmentally exposed children in Mexico City. Pediatrics. 2006 Aug;118(2):e323-30. <u>Other child development tests used:</u> N/A

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Bayley Scales of Infant Development (BSID)	II	Mexico	infants 12.5 - 23.5 months	locally appropriate Spanish	"To explore anthropometric indicators and mental development in very-low-income children in the second year of life. "	"There was a significant decline in HAZ and in age-adjusted MDI score across the second year of life. Although the children showed MDI scores close to the mean, normed US values at 13-14 months, the scores were significantly lower than expected in older children ($P < 0.0001$), even after controlling for socio-economic status and parental characteristics. At 13-14 months, only 3% of children received scores below 70 (less than minus two standard deviations), whereas by 19-20 months, almost 17% of children were performing below this level. No socio-economic or parental characteristics were significant predictors of HAZ or MDI. Parallel deficits are evident in both height-for-age and cognitive functioning during the second year of life in low-income Mexican infants. The consistency of these growth and development findings further stresses the need for targeted interventions to reduce the vulnerability of low-income Mexican children very early in life."	Fernald LC, Neufeld LM, Barton LR, Schnaas L, Rivera J, Gertler PJ.	2006	Fernald LC, Neufeld LM, Barton LR, Schnaas L, Rivera J, Gertler PJ. Parallel deficits in linear growth and mental development in low-income Mexican infants in the second year of life. <i>Public Health Nutr.</i> 2006 Apr;9(2):178-86. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Nicaragua	8.4 months (mean age)	Spanish	Physically abused pr neglected children were evaluated with the Mental Scale of the Bayley Scales of Infant Development and reevaluated subsequently.	The study did not detect significant cognitive deficiencies in these children; if we intervene early enough with abused children, they will have adequate cognitive development.	Mirabal et al.	1990	Mirabal B, De Jesus H. Early detection and cognitive rescue of abused children. <i>Bol Asoc Med P R.</i> 1990 May;82(5):200-3. Spanish.
Bayley Scales of Infant Development (BSID)	II	Nigeria	Eight different age groups between 8 weeks and 30 months.	Not specified	Study set out to establish normative data for psychomotor development on Nigerian children.	The subjects scored above the normal values in the five fields of development for each of the instruments. Furthermore, from the normative data derived, there were high scores in the motor and personal -social fields, especially among subjects in the lower age groups of 8,16,and 32 weeks when compared to data from the western world.	Aina et al.	2005	Aina OF, Morakinyo O. Normative data on mental and motor development in Nigerian children. <i>West Afr J Med.</i> 2005 Apr-Jun;24(2):151-6.
Bayley Scales of Infant Development (BSID)	II	Phillipines	6 to 36 months	Not specified	Study to assess verbal and manipulative skills of children tested in a Phillipines mountain village.	Children with with histories of severe malnutrition obtained very low scores. Several subjects were too sick to respond at all. Because the toddlers can get all they want by whimpering, they have little opportunity or need to acquire self-supporting verbal or manipulative skills. Furthermore, generally poor sanitation predisposes to a high prevalence of the diarrhea-malabsorption-malnutrition sequence, which exacerbates already existing dietary deficiencies.	Guthrie et al.	1976	Guthrie, George M; Masangkay, Zenaida; Guthrie, Helen A Behavior, Malnutrition, and Mental Development. <i>Journal of Cross-Cultural Psychology</i> , 1976, 7, 2, Jun, 169-180.
Bayley Scales of Infant Development (BSID)	II	Poland	infants (born 33-43 weeks of gestation) 0-36 months, BSID-II administered	Polish	"Purpose of the study was to describe the usual pattern of fish consumption during pregnancy in Poland and explain the variability of	"Self-reported weekly amount of fish consumption during the first two trimesters of pregnancy correlated positively with umbilical cord mercury concentrations ($r(s)=0.22$, $p<0.0001$). The corresponding correlation coefficient for the fish	Jedrychowski W, Perera F, Jankowski J, Rauh V, Flak E, Caldwell KL,	2007	Jedrychowski W, Perera F, Jankowski J, Rauh V, Flak E, Caldwell KL, Jones RL, Pac A, Lisowska-Miszczuk I. Fish consumption in pregnancy,

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
			at 12, 24, and 36 months		prenatal mercury exposure due to fish intake by mothers. The other endpoint of the study was the assessment of the cognitive and psychomotor development of infants related to prenatal mercury exposure over the 3-year follow-up."	consumption in the third trimester of pregnancy was 0.21, $p < 0.0001$. There was an inverse association between mercury exposure and both MDI (beta regression coeff. = -2.8, $p = 0.01$) and PDI scores (beta coeff. = -2.3, $p = 0.04$) at 12 months of age. Subsequent BSID-II testing at 24 and 36 months did not confirm significant association between exposure and cognitive or psychomotor function. The estimates of association between mercury prenatal exposure and the development of infants, which were based on the longitudinal analysis of all BSID-II measurements done in the follow-up (generalized estimating equations statistical model) showed that the performance deficit observed at 12 months of age was of border significance."	Jones RL, Pac A, Lisowska-Miszczuk I		cord blood mercury level and cognitive and psychomotor development of infants followed over the first three years of life: Krakow epidemiologic study. Environ Int. 2007 Nov;33(8):1057-62. Epub 2007 Jul 23 <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Poland	12 month old infants	Not specified	"The aim of the study is to assess the cognitive and psychomotor status of 1-year-old infants whose mothers were exposed to low, but varying, amounts of mercury during pregnancy."	"The geometric mean (GM) for maternal blood mercury level for the group of infants with normal neurocognitive performance was lower (GM = 0.52 mug/L; 95% confidence interval [CI], 0.46-0.58) than that observed in the group with delayed performance (GM = 0.75 mug/L; 95% CI, 0.59-0.94), and this difference was significant ($p = 0.010$). The GM of cord blood mercury level in the normal group also was lower (GM = 0.85 mug/L; 95% CI, 0.78-0.93) than that observed in the group with delayed performance (GM = 1.05 mug/L; 95% CI, 0.87-1.27), and this difference was of borderline significance ($p = 0.070$). The relative risk (RR) for delayed performance increased more than threefold (RR = 3.58; 95% CI, 1.40-9.14) if cord blood mercury level was greater than 0.80 mug/L. Risk for delayed performance in the group of infants with greater maternal mercury levels (> 0.50 mug/L) also was significantly greater (RR = 2.82; 95% CI, 1.17-6.79) compared with children whose mothers had mercury levels less than 0.50 mug/L. The results may be of public health importance because delayed psychomotor or mental performance in infants is assumed to be an indicator of later neurocognitive development in children, which may persist into adult life."	Jedrychowski W, Jankowski J, Flak E, Skarupa A, Mroz E, Sochacka-Tatara E, Lisowska-Miszczuk I, Szpanowska-Wohn A, Rauh V, Skolicki Z, Kaim I, Perera F.	2006	Jedrychowski W, Jankowski J, Flak E, Skarupa A, Mroz E, Sochacka-Tatara E, Lisowska-Miszczuk I, Szpanowska-Wohn A, Rauh V, Skolicki Z, Kaim I, Perera F. Effects of prenatal exposure to mercury on cognitive Ann Epidemiol. 2006 Jun;16(6):439-47. Epub 2005 Nov 7. and psychomotor function in one-year-old infants: epidemiologic cohort study in Poland. Ann Epidemiol. 2006 Jun;16(6):439-47. Epub 2005 Nov 7. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Portugal	24 months	Portuguese	Study to compare the mental and motor development at 24 months of 152 full-term infants, born with low (<2500g) and appropriate birth weight (3000 to 3499g).	The infants born with low weight had on average significantly lower mental and motor indexes than those born with appropriate weight ($p < 0.001$), with a difference of 9.1 and 10.2 points, respectively. But socioeconomic conditions and environmental stimulation explained 23% and 21%, respectively, of the variation in these indexes.	Eickmann et al.	2002	Eickmann SH, Lira PI, Lima MC. Mental and motor development at 24 months of full-term low birthweight infants. Arq Neuropsiquiatr. 2002 Sep;60(3-B):748-54. Portuguese.
Bayley Scales	II	Romania	infants 5-31	Not specified	"We assess individual	"Children raised in institutions demonstrated	Smyke AT,	2007	Smyke AT, Koga SF, Johnson

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
of Infant Development (BSID)			months	in abstract or methods	differences in the caregiving environments of young children being raised in institutions in Romania in relation to developmental characteristics such as physical growth, cognitive development, emotional expression, and problem and competence behaviors."	marked delays in cognitive development, poorer physical growth, and marked deficits in competence. Individual differences in caregiving environment were associated with cognitive development, competence, and negative behavior among these young children being reared in institutions. These data confirm previous findings regarding deficits associated with institutional care and extend our understanding of the impact of individual differences in caregiving quality on the development of young children in institutions."	Koga SF, Johnson DE, Fox NA, Marshall PJ, Nelson CA, Zeanah CH; BEIP Core Group.		DE, Fox NA, Marshall PJ, Nelson CA, Zeanah CH; BEIP Core Group. The caregiving context in institution-reared and family-reared infants and toddlers in Romania. <i>J Child Psychol Psychiatry</i> . 2007 Feb;48(2):210-8. <u>Other child development tests used:</u> Laboratory Assessment of Temperment Battery
Bayley Scales of Infant Development (BSID)	II	Seychelles	infants 0-30 months, BSID-II administered at 9 and 30 months	Not specified in abstract or methods	"The present study tested the hypothesis that the intake of selected nutrients in fish or measures of maternal nutritional status may represent important confounders when estimating the effects of prenatal methylmercury exposure on child development."	"The primary analysis examined the associations between MeHg, maternal nutritional measures and children's scores on the BSID-II and showed an adverse association between MeHg and the mean Psychomotor Developmental Index (PDI) score at 30 months. Secondary analyses of the association between the PDI and only MeHg alone or nutritional factors alone showed only a borderline significant association between MeHg and the PDI at 30 months and no associations with nutritional factors. One experimental measure at 5 months of age was positively associated with iodine status, but not prenatal MeHg exposure. These findings suggest a possible confounding role of maternal nutrition in studies examining associations between prenatal MeHg exposures and developmental outcomes in children."	Davidson PW, Strain JJ, Myers GJ, Thurston SW, Bonham MP, Shamlaye CF, Stokes-Riner A, Wallace JM, Robson PJ, Duffy EM, Georger LA, Sloane-Reeves J, Cernichiari E, Canfield RL, Cox C, Huang LS, Janciuras J, Clarkson TW	2008	Davidson PW, Strain JJ, Myers GJ, Thurston SW, Bonham MP, Shamlaye CF, Stokes-Riner A, Wallace JM, Robson PJ, Duffy EM, Georger LA, Sloane-Reeves J, Cernichiari E, Canfield RL, Cox C, Huang LS, Janciuras J, Clarkson TW. Neurodevelopmental effects of maternal nutritional status and exposure to methylmercury from eating fish during pregnancy. <i>Neurotoxicology</i> . 2008 Sep;29(5):767-75. Epub 2008 Jun 11 <u>Other child development tests used:</u> Fagan Infantest (FTII), Visual Expectation Paradigm (VEXP), A-not-B, Delayed Spatial Alteration (DSA)
Bayley Scales of Infant Development (BSID)	II	South Africa	infants 18-30 months	Not specified in abstract	"The aim of this study was to determine the extent of delay in acquisition of language, cognitive, and motor skills of children infected with human immunodeficiency virus (HIV)."	"Mean cognitive development was 7.63 months delayed (P < 0.001) and mean motor development was 9.65 months delayed (P < 0.001), with 97.5% of the sample functioning below expected motor and cognitive age. Eighty-five percent of the sample demonstrated gross motor delay, which was the most adversely affected skill. Language was descriptively analyzed, revealing global language delay in 82.5% of the children. Gross motor delay may be attributed to decreased strength or to HIV encephalopathy. Cognitive delay may be because of disease progression and structural damage to the brain, and language delay may be attributed to neurological impairment, cognitive delay, or	Baillieu N, Potterton J	2008	Baillieu N, Potterton J. The extent of delay of language, motor, and cognitive development in HIV-positive infants. <i>J Neurol Phys Ther</i> . 2008 Sep;32(3):118-21 <u>Other child development tests used:</u> N/A

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						environmental deprivation. Based on these results, infants and children should be referred to physiotherapy to address the gross motor deficits noted."			
Bayley Scales of Infant Development (BSID)	II	Taiwan	Not specified	Chinese	Used to investigate the applicability of BSID-II in diagnosing children with developmental delay in Kaohsiung area.	Highly reliable for child development assessment. A special norm for developmentally delayed children and quality control of examiners are suggested.	Huang et al.	2000	Huang HL, Chuang SF, Jong YJ, Yu L, Shieh YL. Applicability of BSID-II in diagnosing developmental delay at Kaohsiung area. Kaohsiung J Med Sci. 2000 Apr;16(4):197-202.
Bayley Scales of Infant Development (BSID)	II	Taiwan	infants 6-24 months, BSID-II administered at 6, 12, 18, and 24 months	Not specified in abstract	"To establish the normative data of the Bayley Scales of Infant Development-Second Edition (BSID-II) on Taiwanese infants from age 6 to 24 months and to explore the factors that relate to their mental and motor development."	"Taiwanese infants' Bayley mental and motor raw scores were lower than the United States norms from age 6 to 24 months, however, the discrepancy gradually declined with increasing age. Gender, intrauterine growth status, birth order, region of residence, maternal education, and paternal occupation were shown to have longitudinal associations with their mental and/or motor scores. Differences existed in the mental and motor development among Taiwanese and American infants. Our preliminary norms of the BSID-II may be more appropriate than the United States norms for Taiwanese children."	Wu YT, Tsou KI, Hsu CH, Fang LJ, Yao G, Jeng SF; Taiwan Infant Developmental Collaborative Study Group	2008 (Epub 2007)	Wu YT, Tsou KI, Hsu CH, Fang LJ, Yao G, Jeng SF; Taiwan Infant Developmental Collaborative Study Group. Brief report: Taiwanese infants' mental and motor development-6-24 months. J Pediatr Psychol. 2008 Jan-Feb;33(1):102-8. Epub 2007 Aug 21 <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Taiwan	infants at 6, 12, 18, and 24 months.	Not specified in abstract or methods	"To determine the developmental trajectories of very-low-birthweight (VLBW) infants during the first 2 years of life, and investigate the most contributory predictors of diverse trajectories."	"Five trajectories of cognitive development based on the mental developmental indices from the age of 6 to 24 months were determined, including average-stable (group A, 20.1%), average-decline to borderline delay (group B, 34%), borderline delay-catch-up to average (group C, 20.2%), borderline delay-decline to significant delay (group D, 17.2%) and significant delay-stable (group E, 8.5%). Using group A as the reference category, we determined 6-month neurological status and maternal education as the most significant predictors for various trajectories ($p < 0.01$). Infants with transient or definite neurological abnormality and/or low maternal education had higher odds of displaying the disadvantageous trajectories (group B, D and E; odds ratios, 1.79-46.4). VLBW infants with neurological abnormalities and/or low maternal education had high risk of developmental decline and might benefit from early intervention"	Wang LW, Wang ST, Huang CC.	2008	Wang LW, Wang ST, Huang CC. Preterm infants of educated mothers have better outcome. Acta Paediatr. 2008 May;97(5):568-73. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Taiwan	full-term infants ages 6-18 months	Not specified in abstract	"This study investigated the concurrent validity of the Comprehensive Developmental Inventory	"The results showed that correlation coefficients for Developmental Ages between both tests on cognitive and motor subtests were high ($r = .91-.95$) and for Developmental Quotients were moderate (r	Liao HF, Yao G, Wang TM	2008	Liao HF, Yao G, Wang TM. Concurrent validity in Taiwan of the Comprehensive Developmental Inventory for

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					for Infants and Toddlers (CDIIT) with the Bayley Scales of Infant Development-II (BSID-II) in full-term infants. "	= .57-.67). Moderate classification agreement was found in the two scales (quadratic weighted kappa = .50-.53). Developmental Quotients classification for the CDIIT tended to be a little higher than for the BSID-II. It was concluded that although acceptable concurrent validity was found for the Motor and Cognitive subtests of the CDIIT, the tester should be cautious to compare Developmental Quotients obtained from the above two tests in clinical or in research settings."			Infants and Toddlers who were full-term infants. Percept Mot Skills. 2008 Aug;107(1):29-44. <u>Other child development tests used:</u> Comprehensive Developmental Inventory for Infants and Toddlers (CDIIT)
Bayley Scales of Infant Development (BSID)	II	Taiwan	infants 6-24 months	Not specified	"The objective of this study is to evaluate the physical and mental development of infants born on in-vitro maturation (IVM) programs."	"We found all the children with normal karyotype and without major malformation in both IVM and non-IVM groups. The mean Mental Development Index scores for IVM subjects and the comparison group were 92.71+/-10.47 and 97.19+/-8.88, respectively (p=0.074). The mean Psychomotor Development Index scores were 96.67+/-8.91 and 96.19+/-7.05, respectively (p=0.817). This is the first study designed to evaluate the physical growth and developmental indices of IVM children with combinational priming protocol of FSH and hCG. Our results suggest that IVM children didn't show developmental delay during infancy and early childhood."	Shu-Chi M, Jiann-Loung H, Yu-Hung L, Tseng-Chen S, Ming-I L, Tsu-Fuh Y.	2006	Shu-Chi M, Jiann-Loung H, Yu-Hung L, Tseng-Chen S, Ming-I L, Tsu-Fuh Y. Growth and development of children conceived by in-vitro maturation of human oocytes. Early Hum Dev. 2006 Oct;82(10):677-82. Epub 2006 May 11. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Taiwan	6-18 months	Not specified in abstract or methods.	"To develop an efficient developmental screening instrument for interview purposes."	"In testing the concurrent validity, the structural equation model showed that Taiwan Birth Cohort Study dimensions correlated with the Bayley Scale of Infant Development dimensions. In testing the predictive validity of the Taiwan Birth Cohort Study, the correlation between the 6- and 18-month results of the motor index was 0.29, and the language social index was 0.31. The Taiwan Birth Cohort Study 6- and 18-month scales fulfill the criteria of developmental norm, predictive validity, validity and identification of relative risk for screening instruments. As a self-report checklist, it is economical and efficient, thus, it can be utilized in clinical and community settings."	Lung FW, Shu BC, Chiang TL, Lin SJ.	2008	Lung FW, Shu BC, Chiang TL, Lin SJ. Efficient developmental screening instrument for 6- and 18-month-old children in the Taiwan Birth Cohort Pilot Study. Acta Paediatr. 2008 Aug;97(8):1093-8. Epub 2008 May 7 <u>Other child development tests used:</u> Taiwan Birth Cohort Study scale
Bayley Scales of Infant Development (BSID)	I or II (not indicated in abstract)	Taiwan(China in World Bank classification)	children at 6, 12, 18, and 24 months	Not specified in abstract	"The specific aim of our study was to evaluate whether there is significant difference in the Bayley developmental index scores at 6, 12, 18 and 24 months of corrected age for very-low-birthweight (birth body weight <1500 gm, VLBW) infants with or without placental CAM."	"We found that 56.8% of placentas presented a picture of CAM. In comparison of the neonatal characteristics, VLBW infants with CAM had shorter gestational age (27.9 +/- 2.8 vs. 30.0 +/- 3.7 weeks, p = 0.003), lower Cesarean delivery rate (48.1% vs. 73.2%, p = 0.011), more maternal steroid use (44.4% vs. 12.2%, p = 0.004) and higher incidence of preterm premature rupture of membrane (PPROM, 37.0% vs. 12.2%, p = 0.009). In comparison of neonatal outcomes, the CAM group had higher incidence of bronchopulmonary	Mu SC, Lin CH, Sung TC, Chen YL, Lin YC, Lee CC, Chen TJ, Lin MI, Jow GM.	2007	Mu SC, Lin CH, Sung TC, Chen YL, Lin YC, Lee CC, Chen TJ, Lin MI, Jow GM. Neurodevelopmental outcome of very-low-birth-weight infants with chorioamnionitis. Acta Paediatr Taiwan. 2007 Jul-Aug;48(4):207-12. <u>Other child development tests used:</u>

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						dysplasia (BPD, 40.7% vs. 19.5%, p = 0.044), more mechanical ventilation (87.0% vs. 27/41, p = 0.023) and intubation (68.5% vs. 46.3%, p = 0.049), and more median days of ventilation (23.1 +/- 29.1 vs. 7.8. +/- 13.7 days, p = 0.001). As for the follow-up, at any test age, either the mean (Mental Development Index (MDI) / (Psychomotor Development Index (PDI) scores of Bayley test or the incidence of score below 85, there was no significant difference in both groups. The VLBW infants with histologic chorioamnionitis were not associated with an increased risk of lower MDI or PDI scores at the corrected ages of 6, 12, 18 and 24 months compared with the non-CAM control group."			N/A
Bayley Scales of Infant Development (BSID)	II	Taiwan (China in World Bank classification)	infants at 6 and 12 months	Not specified in abstract, but scored using Taiwanese norms	"This study examined the developmental and clinical outcomes in very-low-birthweight (VLBW; < or =1500g) infants with and without bronchopulmonary dysplasia (BPD) throughout infancy, and assessed if BPD predicted poor developmental outcome beyond the effects of other risk factors."	"Compared with infants without BPD, infants with BPD had higher rates of clinical morbidity, and those with severe BPD further exhibited higher incidences of developmental delay throughout infancy. BPD predicts poor 1-year developmental and clinical outcomes in VLBW infants for which effects are well correlated to the NIH consensus definition."	Jeng SF, Hsu CH, Tsao PN, Chou HC, Lee WT, Kao HA, Hung HY, Chang JH, Chiu NC, Hsieh WS.	2008	Jeng SF, Hsu CH, Tsao PN, Chou HC, Lee WT, Kao HA, Hung HY, Chang JH, Chiu NC, Hsieh WS. Bronchopulmonary dysplasia predicts adverse developmental and clinical outcomes in very-low-birthweight infants. Dev Med Child Neurol. 2008 Jan;50(1):51-7. <u>Other child development tests used:</u> Neonatal Neurobehavioral Examination - Chinese version
Bayley Scales of Infant Development (BSID)	II	Taiwan +C252	children at 6, 12, 18, and 24 months	Not specified in abstract or methods	"The aim of this study was to investigate the neurodevelopmental outcome in very low birth weight infants with postnatal subependymal cysts."	"Preterm infants with postnatal subependymal cysts had a significantly lower Psychomotor Development Index (P = .034) and were more likely than the normal group to have motor developmental delay (Psychomotor Development Index <70) (P = .013). The findings indicate that postnatal subependymal cyst is a significant risk factor for impaired motor development in very low birth weight infants (odds ratio 5.73, 95% confidence interval 1.57-20.97)."	Chuang YC, Lee C, Chiu NC, Shu CH, Hung HY, Kao HA, Chang JH.	2007	Chuang YC, Lee C, Chiu NC, Shu CH, Hung HY, Kao HA, Chang JH. Neurodevelopment in very low birth weight premature infants with postnatal subependymal cysts. J Child Neurol. 2007 Apr;22(4):402-5. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Tanzania	6 to 18 months	Kiswahili	To determine the association between maternal multivitamin supplementation and the mental and psychomotor development of children who are born to HIV-1-	Multivitamin supplementation was associated significantly with a mean increase in Psychomotor Development Index score of 2.6 (95% confidence interval: 0.1-5.1). Multivitamins were also significantly protective against the risk for developmental delay on the motor scale (relative risk: 0.4; 95% confidence interval: 0.2-0.7) but not	McGrath et al.	2006	McGrath N, Bellinger D, Robins J, Msamanga GL, Tronick E, Fawzi WW. Effect of maternal multivitamin supplementation on the mental and psychomotor development of children who are born to

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					infected mothers in Tanzania.	on the Mental Development Index. Vitamin A supplementation had no significant effect on these outcomes. Maternal multivitamin supplements provide a low-cost intervention to reduce the risk for developmental delays among infants who are born to HIV-positive mothers in developing countries.			HIV-1-infected mothers in Tanzania. Pediatrics. 2006 Feb;117(2):e216-25
Bayley Scales of Infant Development (BSID)	II	Tanzania	6-18 months, administered at 6,12, 18 months	Kiswahili translation of all instructions	"To determine the association between maternal multivitamin supplementation and the mental and psychomotor development of children who are born to HIV-1-infected mothers in Tanzania, as secondary endpoints in a randomized trial that investigated the effect of maternal multivitamin supplementation on HIV-1 vertical transmission and progression. "	"Multivitamin supplementation was associated significantly with a mean increase in Psychomotor Development Index score of 2.6 (95% confidence interval: 0.1-5.1). Multivitamins were also significantly protective against the risk for developmental delay on the motor scale (relative risk: 0.4; 95% confidence interval: 0.2-0.7) but not on the Mental Development Index. Vitamin A supplementation had no significant effect on these outcomes. Maternal multivitamin supplements provide a low-cost intervention to reduce the risk for developmental delays among infants who are born to HIV-positive mothers in developing countries."	McGrath N, Bellinger D, Robins J, Msamanga GI, Tronick E, Fawzi WW.	2006	McGrath N, Bellinger D, Robins J, Msamanga GI, Tronick E, Fawzi WW. Effect of maternal multivitamin supplementation on the mental and psychomotor development of children who are born to HIV-1-infected mothers in Tanzania. Pediatrics. 2006 Feb;117(2):e216-25. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development (BSID)	II	Tanzania	6-18 months, administered at 6,12, 18 months	Kiswahili translation of all instructions and culturally appropriate modifications	"To determine the association between the timing of mother-to-child transmission of human immunodeficiency virus (HIV)-1 and neurodevelopment among children born to HIV-1 infected mothers in Tanzania."	"Children who tested HIV-1-positive at birth had significantly higher decreases per month in MDI and PDI than HIV-1-negative children; 1.1 [95% confidence interval (95% CI) 0.4, 1.8] for MDI and 1.4 (95% CI 0.0, 2.7] for PDI. Children who tested HIV-1-positive after birth had an additional 0.6 (95% CI 0.1, 1.1) point decrease in MDI per month and a 0.6 (95% CI 0.0, 1.1) higher decrease in PDI each month than HIV-1-negative children. Testing HIV-1-positive at birth was associated with a 14.9 (95% CI 5.0, 44.7) times higher rate of becoming developmentally delayed in mental function, while testing HIV-1-positive after birth was associated with a 3.2 (95% CI 1.6, 6.4) times higher rate than in uninfected children. HIV-1 infected infants performed worse on tests of neurodevelopment and were significantly more likely to be identified as developmentally delayed in the first 18 months of life than HIV-1-negative children. The effect of HIV-1 infection on neurodevelopment scores and the risk of developmental delay may be highest among those who are already HIV-1 infected at birth."	McGrath N, Fawzi WW, Bellinger D, Robins J, Msamanga GI, Manji K, Tronick E.	2006	McGrath N, Fawzi WW, Bellinger D, Robins J, Msamanga GI, Manji K, Tronick E. The timing of mother-to-child transmission of human immunodeficiency virus infection and the neurodevelopment of children in Tanzania. Pediatr Infect Dis J. 2006 Jan;25(1):47-52. <u>Other child development tests used:</u> N/A
Bayley Scales of Infant Development	II	Turkey	Children 6-30 months old with a three month	Turkish	Study to assess the effects of iron deficiency and iron treatment on developmental	Iron deficiency may cause lower mental and motor test scores in infants and these adverse effects can be improved by iron therapy.	Akman et al.	2004	Akman M, Cebeci D, Okur V, Angin H, Abali O, Akman AC. The effects of iron deficiency

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
(BSID)			follow up.		test scores in infants.				on infants' developmental test performance. <i>Acta Paediatr.</i> 2004 Oct;93(10):1391-6.
Bayley Scales of Infant Development (BSID)	II	Turkey	Children 6-30 months old with a three month follow up.	Turkish	Study to assess the effects of iron deficiency and iron treatment on developmental test scores in infants.	Iron deficiency may cause lower mental and motor test scores in infants and these adverse effects can be improved by iron therapy.	Akman et al.	2004	Akman M, Cebeci D, Okur V, Angin H, Abali O, Akman AC. The effects of iron deficiency on infants' developmental test performance. <i>Acta Paediatr.</i> 2004 Oct;93(10):1391-6.
Bayley Scales of Infant Development (BSID)	II	Turkey	6 months and 9 months	Turkish	A study to examine the effect of iron supplementation on performance in the BSID and anthropometric measurement in 6-month-old iron-sufficient healthy infants.	The BSID scores of infants in both groups (iron supplementation and no iron supplementation) were within the normal range on admission and at the end of the study period.	Yalcin et al.	2000	Yalcin SS, Yurdakok K, Acikgoz D, Ozmert E. Short-term developmental outcome of iron prophylaxis in infants. <i>Pediatr Int.</i> 2000 Dec;42(6):625-30.
Bayley Scales of Infant Development (BSID)	I	Turkey	3-24 months, administered every 3 months from birth	Not specified in abstract or methods	"The objective of the present study was to evaluate the term newborn infants admitted to Gazi University Hospital neonatal intensive care unit (NICU) for hypernatremic dehydration between June 2001 and June 2003 and compare the results with those of the literature search performed via MEDLINE for infants with the same diagnosis."	"Between June 2001 and June 2003, 28 newborns were admitted to NICU with hypernatremic dehydration. Literature review found 178 newborns with the same diagnosis since 1979 and detailed information was available for 150 patients. In the study and MEDLINE groups, respectively mean days of admission were 3.39 and 11.7; mean serum sodium, 156.5 and 178.6 mEq/L; and mean % weight loss, 11.5 and 25.7. Long-term follow up was performed in 15 patients from the study group, and two patients were found to be severely neurologically delayed. Five patients were found to have moderate risk scores by BINS tests. Early follow-up visits of newborns soon after discharge are important to determine risk for hypernatremic dehydration. Long-term follow up of this group of babies is also required to gather knowledge about the repercussions of early hypernatremic dehydration later in life."	Ergenekon E, Unal S, Gücüyener K, Soysal SE, Koç E, Okumus N, Türkyilmaz C, Onal E, Atalay Y.	2007	Ergenekon E, Unal S, Gücüyener K, Soysal SE, Koç E, Okumus N, Türkyilmaz C, Onal E, Atalay Y. Hypernatremic dehydration in the newborn period and long-term follow up. <i>Pediatr Int.</i> 2007 Feb;49(1):19-23. <u>Other child development tests used:</u> Bayley Infant Neurodevelopmental Screener (BINS)
British Ability Scales	II	India	Preschool children ; average age 5.3 months of age	English	This paper reports the findings of a study investigating factors associated with achievement and progress in young Indian children learning to read in English. A longitudinal design was used in which children's reading was followed-up from the end of nursery for 17 months.	"Word recognition at the end of nursery consistently emerged as the variable most strongly associated with progress in reading. While this study provides further support for word recognition as a factor strongly associated with reading achievement and progress, the results need to be interpreted in the context of the children's generally poor scores on comprehension of text. This has serious implications for teaching practices that emphasise word recognition without laying stress on language comprehension, particularly in relation to bilingual children whose education is in a language that is not	Sen, R., & Blatchford, P	2001	Sen, R., & Blatchford, P. (2001, June). Reading in a second language: Factors associated with progress in young children. <i>Educational Psychology, 21</i> (2), 189-202. <u>Other child development tests used:</u> Two performance items, the Mazes and Picture Completion, from the Wechsler Pre-school

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						the mother tongue."			and Primary Scales of Intelligence were included as tests of logical and analytical thinking. Neale Analysis of Reading Ability (Neale, 1989)
British Ability Scales	II	Zimbabwe	N/A				Mpofu, E.	1995	Mpofu, E. (1995, February). Antecedents of children's performance on class inclusion tasks: Some Zimbabwean evidence. <i>International Journal of Psychology</i> , 30(1), 19-33.
British Ability Scales	II	Kenya, Uganda, Zanzibar	pre-school	Local languages as appropriate	To assess the impact of the Madrasa program on the cognitive development of children and the teaching and learning environment of preschools	"Overall, the MRC preschools provide the greatest value added in terms of children's intellectual development. Children from MRC preschools did significantly better than either those who did not attend preschool or those who attended other preschools in East Africa. MRC preschool children had a margin of 42 percent higher value-added mean scores against the home children. Also MRC preschool children later perform better in school, as compared with other children from home or from normative preschools. Children from the MRC preschool programs were more likely to be ranked in the upper half of their class on examinations, compared with children from home (with no preschool experience) or those from other preschools."	Garcia M, Pence A, Evans JL (Eds).	2008	Garcia, M., Pence, A., & Evans, J. L. (Eds.). (2008). <i>Africa's Future, Africa's Challenge: Early Childhood Care and Development in Sub-Saharan Africa</i> . Washington, DC: The International Bank for Reconstruction and Development. <u>Other child development tests used:</u> African Child Test
Cambodian Developmental Assessment Test	N/A	Cambodia	3-5 yr	English and Cambodian	Compare three different early childhood education programs and a control group on child development outcomes	All groups with preschool scored higher than the control group, formal preschools had higher scores than two informal groups who did not differ from each other on child outcomes.	Rao, N., & Pearson, E. With V. Pearson & M. Conostas (2007).	2007	Rao, N., & Pearson, E. With V. Pearson & M. Conostas (2007). <i>An Evaluation of Early Childhood Education and Care Programs in Cambodia</i> . Report to UNICEF. Cambodia.
Cambodian Developmental Assessment Test	N/A	Cambodia	3 to 6 years	English and Cambodian	To examine the effectiveness of various Early Childhood Care and Education programs in Cambodia.	Children who participated in early childhood programmes (State Preschool, Community Preschool and Home-based Programmes) had significantly better developmental functioning than children in the control group.	UNICEF Cambodia	2007	Rao N. & Pearson E. (2007). <i>An evaluation of Early Childhood Care and Education Programmes in Cambodia</i> . Unpublished manuscript.
Delay of Gratification	N/A	China	preschool children	Chinese	Examination of young children's abilities to delay gratification by following experimenters' instructions to avoid playing with a highly attractive novel toy.	Results indicated that high motivation produced less noncompliant behavior, while high stress produced more noncompliant behavior. Girls exhibited less noncompliant behavior than did boys. Implications were discussed for factors contributing to the prosocial behavior of self-control.	Wang et al	2003	Wang A, Karns J, Meredith W. Motivation, Stress, Self-Control Ability, and Self-Control Behavior of Preschool Children in China. <i>Journal of Research in Childhood Education</i> , 2003, 17(2), Spr-Sum 2003. pp. 175-187
Delay of	N/A	China and	3 to 5 year olds	English and	Cross-cultural comparison	Chinese and Australian children used different types	Lizhu et al.	2005	Lizhu L, Jiangyang W, Wen L,

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Gratification		Australia		Chinese	of delay of gratification strategies and performance in preschool-aged children.	of strategies for dealing with the delay. Overall results indicated that cultural differences influence children's performance on delay of gratification tasks.			Cuskelly M, Zhang A. Strategies used by 3 to 5 years old children on a self-imposed delay of gratification task including a cross-cultural comparison between China and Australia. Acta Psychologica Sinica, 2005, 37(2), 224-232.
Delay of Gratification	N/A	Japan	4.2 years to 6.7 years	Japanese	Studied developmental changes in <i>delay</i> -choice and <i>delay</i> -maintenance behavior. Human subjects: 90 normal male and female Japanese children (4 yrs 2 mo to 6 yrs 8 mo). Ss were assigned to situations in which the difference in value between immediate and delayed reward was small or large		Mitsutomi	1988	Mitsutomi M. Development of delay of gratification in preschool children, 1988, Japanese Journal of Psychology, 59(1), 57-60.
Denver Developmental Materials	II	Brazil	various ages (review article)	various studies (review article), informal cultural adaptation for Brazil	"To review the literature regarding screening psychomotor tests for the early identification of developmental problems."	"A total of 455 references were listed, and 174 studies were selected for this review based on title, relevance, and abstract. Only original and electronically available material, from 1985 forward, with information on design, applicability, and psychometric properties of those tests was included. Screening tests are important to speed the beginning of treatment measures in order to allow for better developmental outcome. Among the many tests that can be employed for this purpose, the DENVER II and the Alberta Infant Motor Scale are the most often used in Brazilian studies. The Movement Assessment of Infants is starting to be used in our country. Two other tests are recommended in the literature due to their high sensibility and specificity: the Test of Infant Motor Performance and the General Movements."	Santos RS, Araújo AP, Porto MA.	2008	Santos RS, Araújo AP, Porto MA. Early diagnosis of abnormal development of preterm newborns: assessment instruments. J Pediatr (Rio J). 2008 Jul-Aug;84(4):289-99. <u>Other child development tests used:</u> Bayley and other
Denver Developmental Materials	N/A	Armenia	Not specified	Armenian	Study to establish the cultural validity of the Denver Developmental Screening Test norms for Armenian children.	Established norms are valid for children of Armenian culture. Language, cultural differences, and child rearing practices contributed to deviations from the norms.	Akaragian et al.	1992	Akaragian, S. and C. Dewa. Standardization of the Denver Developmental Screening Test for Armenian children. J Pediatr Nurs. 1992;7(2): 106-9.
Denver Developmental Materials	N/A	Brazil	11-72 months	Not specified	Study to detect neurological abnormalities in human immunodeficiency virus (HIV) infected children.	On the DDST, group I (HIV-Infected) presented significantly more failures than group II (HIV seroreverters. HIV-infected children a neurodevelopment delay occurs early in the disease, and it can be detected by screening tests.	Bruck et al.	2001	Bruck, I., T. T. Tahan, et al. Developmental milestones of vertically HIV infected and seroreverters children: follow up of 83 children. Arq Neuropsiquiatr. 2001; 59(3-B): 691-5.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Denver Developmental Materials	N/A	Brazil	6-59 months	Not specified	Concurrent validity of maternal opinion of child development was estimated in a cross-sectional, population-based survey of children using a standard measure devised from the Denver Developmental Screening Test.	Sensitivity, specificity and negative predictive value increased with maternal education and family income. Positive predictive value was higher in low-income families and children with impairments, low birthweight and long hospital stays. Children at social and clinical risk should be assessed more carefully, even if maternal report is normal or advanced.	de Lourdes.	2005	de Lourdes Drachler M, de Castro Aerts DG et al. Social inequalities in maternal opinion of child development in southern Brazil. <i>Acta Paediatr.</i> 2005; 94(8): 1137-9.
Denver Developmental Materials	Versions II and Developmental Screening Test	Brazil	under 5 year olds	Not specified in abstract	"A method for translating research data from the Denver Test into individual scores of developmental status measured in a continuous scale is presented. "	"Each child's ability age was then estimated by maximum likelihood as the age in this reference population corresponding to the child's success and failures in the test. The score of developmental status is the natural logarithm of this ability age divided by chronological age and thus measures the delay or advance in the child's ability age compared with chronological age. This method estimates development status using both difficulty and discriminating power of each item in the reference population, an advantage over scores based on total number of items correctly performed or failed, which depend on difficulty only. The score corresponds with maternal opinion of child developmental status and with the 3-category scale of the DDST. It shows good construct validity, indicated by symmetrical and homogeneous variability from 3 months upwards, and reasonable results in describing gender differences in development by age, the mean score increasing with socio-economic conditions and diminishing among low-birthweight children. If a standardised measure of development status (z-scores) is required, this can be obtained by dividing the score by its standard deviation. Concurrent and discriminant validity of the score must be examined in further studies."	Drachler Mde L, Marshall T, de Carvalho Leite JC.	2007	Drachler Mde L, Marshall T, de Carvalho Leite JC. A continuous-scale measure of child development for population-based epidemiological surveys: a preliminary study using Item Response Theory for the Denver Test. <i>Paediatr Perinat Epidemiol.</i> 2007 Mar;21(2):138-53. <u>Other child development tests used:</u> N/A
Denver Developmental Materials	Versions II and Developmental Screening Test	Brazil	children 1-5 years (some older than 5 may have been included in the study as well)	Not specified in abstract or methods	"This study evaluated the degree of neurological compromise in HIV-infected children accompanied by the outpatient clinic of infectious diseases and pediatric neurology of the Clinical Hospital of the Federal University of Paraná (UFPR) starting in 1995."	"Neurological and neurodevelopmental alterations were found in 82% of the HIV-infected patients and in 36% of the HIV-seroreverter group (P<0.01). In the CAT/CLAMS test, the development quotient (DQ) of the HIV-infected group was significantly lower than that of the HIV-seroreverter group. CAT/CLAMS scores lower than 70 (mental deficiency) were found in 31% of the HIV-infected patients during the first year of life and in only 1% of the patients of the HIV-seroreverter group, demonstrating the validity of this screening test for precocious detection of alterations in the neurodevelopment of infected patients. The same occurred with the Denver I and II tests, as the HIV-	Tahan TT, Bruck I, Burger M, Cruz CR.	2006	Tahan TT, Bruck I, Burger M, Cruz CR. Neurological profile and neurodevelopment of 88 children infected with HIV and 84 seroreverter children followed from 1995 to 2002. <i>Braz J Infect Dis.</i> 2006 Oct;10(5):322-6. <u>Other child development tests used:</u> CAT/CLAMS

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						infected group failed more frequently than the HIV-seroreverter group. Nine HIV-infected children presented altered brain CT scans; calcification of basal ganglia was the main finding (five cases). Encephalopathy due to HIV causes early arrest of neurodevelopment, which can be detected with screening tests during the first year of life."			
Denver Developmental Materials	II	Brazil	assessed at 12 and 24 months	Not specified in abstract or methods	"To evaluate the influence of gestational age and birth weight on language development and neurodevelopmental outcome at age 3 years in children born preterm."	"NLA children had higher scores on mental and psychomotor ($p<0.01$, $p=0.012$) indexes of Bayley II. Newborns with less than 1500 grams had lower scores on all Bayley scale at age 36 months ($p=0.002$, $p=0.007$ and $p<0.001$). Multivariate analysis suggests an association between gestational age ($p=0.032$), abnormal behavior ($p<0.001$) and delay in language acquisition. Denver test at 12 and 24 months of age was a good predictor of delayed receptive and expressive language at three years of age ($p<0.01$ and $p<0.01$). Children born prematurely with low birth weight had an increased risk of language acquisition delay, and those had also lower cognitive and behavior scores when compared to NLA."	Schirmer CR, Portuguese MW, Nunes ML.	2006	Schirmer CR, Portuguese MW, Nunes ML. Clinical assessment of language development in children at age 3 years that were born preterm. Arq Neuropsiquiatr. 2006 Dec;64(4):926-31. <u>Other child development tests used:</u> Bayley II
Denver Developmental Materials	II	Brazil	12 months old	Not specified in abstract or methods. Version adapted to the Brazilian population	To compare "developmental status at age 12 months according to birth weight and family income" in two Brazilian birth cohorts.	"The prevalence of suspected developmental delay fell from 37.1% in 1993 to 21.4% in 2004 and was inversely proportional to family income and birth weight. Among children born weighing under 2,000 g, there was a fourfold reduction in the prevalence of developmental delay between 1993 and 2004. With regard to family income, the poorest group showed the greatest reduction between the two cohorts--a 30% reduction in risk. Our results confirm the influence of income and birth weight on child development. The decrease in the prevalence of developmental delay in the last decade reflects, among other factors, improvements in neonatal care, increased coverage of developmental monitoring in the first year of life, and longer breastfeeding duration. Despite this reduction, the prevalence of developmental delay is still high, reinforcing the need for early diagnosis and intervention."	Halpern R, Barros AJ, Matijasevich A, Santos IS, Victora CG, Barros FC.	2008	Halpern R, Barros AJ, Matijasevich A, Santos IS, Victora CG, Barros FC. Developmental status at age 12 months according to birth weight and family income: a comparison of two Brazilian birth cohorts. Cad Saude Publica. 2008;24 Suppl 3:S444-50. <u>Other child development tests used:</u> N/A
Denver Developmental Materials	N/A	China (Taipei)	3 to 5 years	Chinese	Tested the DDST-Chinese version	Found that norms need to be adjusted for children in Taipei to account for cultural differences. Children in this sample were especially adept at passing fine motor items.	Chen et al.	2003	Chen CJ, Li IC, Chien LY. Developmental status among 3 to 5-year-old preschool children in three kindergartens in the Peitou District of Taipei City. J Nurs Res. 2003 Jun;11(2):73-81.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Denver Developmental Materials	II	Egypt	assessed at 6, 12, and 18 months	Not specified in abstract	"The purpose of this study was to investigate, in neonates with HIE, the role of early EEG and conventional MRI in the prediction of infants at risk for persistent encephalopathy at 18 months of age."	"Three infants (9%) had mild HIE, 21 infants (62%) had moderate HIE, and 10 infants (29%) had severe HIE. The EEG background was normal, moderately, severely, and extremely discontinuous in eight (24%), three (9%), sixteen (47%), and seven (20%) neonates, respectively. EEG background activities correlated significantly with HIE severity ($p = 0.0001$). MRI findings significantly correlated with EEG background ($p = 0.001$). Normal MRI scans and minimal basal ganglia lesions were always associated with normal EEG background. Patients with severe basal ganglia and thalamic lesions in MRI ($n = 2$) had extreme discontinuous EEG background. For the prediction of poor outcomes, abnormal EEG background activity had a sensitivity (S_n) = 100%, a specificity (S_p) = 100%, positive predictive value (PPV) = 100%, and negative predictive value (NPV) = 100%, whereas values of abnormal MRI scans were S_n of 100%, $S_p = 43%$, PPV = 82%, and NPV=100%. EEG background activity is the best element to predict abnormal outcomes. Severe basal ganglia and thalamic injuries on MRI scans are associated with poor outcomes. Otherwise, MRI does not contribute to the prediction of outcomes at 18 months of age."	El-Ayouty M, Abdel-Hady H, El-Mogy S, Zaghlol H, El-Beltagy M, Aly H.	2007	El-Ayouty M, Abdel-Hady H, El-Mogy S, Zaghlol H, El-Beltagy M, Aly H. Relationship between electroencephalography and magnetic resonance imaging findings after hypoxic-ischemic encephalopathy at term. Am J Perinatol. 2007 Sep;24(8):467-73. <u>Other child development tests used:</u> N/A
Denver Developmental Materials	N/A	Japan	Toddlers	Japanese	Study is to examine the relationship between toddler temperament and developmental delay.	Compared with normal children, for developmentally delayed children: (1) The temperamental category scores of adaptability and persistence were higher, indicating low adaptability and persistence. The prevalence of difficult child, slow to warm up (STWU) child and intermediate high child was relatively higher, with STWU child the highest. (2) The score for the rearing environment was lower.	Asahara et al.	1992	Asahara, K., S. Murashima, et al. Study on the relationship between toddler temperament and development (second report)--the relationship between toddler temperament and developmental delay. Nippon Kosho Eisei Zasshi. 1992; 39(11): 839-47.
Denver Developmental Materials	Versions II and Developmental Screening Test	Malawi	infants and children 0-5 (older than 5 included in the study as well)	Translation and culturally appropriate adaptation of Denver, included re-adaptation and re-translation through piloting	"To create a more culturally relevant developmental assessment tool for use in children in rural Africa."	"An assessment tool with 138 items was created. Face, content and respondent validity was demonstrated. At the consensus meeting, 97% (33/34) of gross motor items were retained in comparison to 51% (18/35) of social items, and 86% (69/80) of items from the Denver II or Denver Developmental Screening Test (DDST) were retained in comparison to 69% (32/46) of the newly created items, many of these having poor reliability and goodness of fit. Gender had an effect on 23% (8/35) of the social items, which were removed. Items not attained by 6 years came entirely from the Denver II fine motor section (4/34). Overall, 110 of the 138 items (80%) were retained in the revised instrument with some items needing further	Gladstone MJ, Lancaster GA, Jones AP, Maleta K, Mtitimila E, Ashorn P, Smyth RL.	2008	Gladstone MJ, Lancaster GA, Jones AP, Maleta K, Mtitimila E, Ashorn P, Smyth RL. Can Western developmental screening tools be modified for use in a rural Malawian setting? Arch Dis Child. 2008 Jan;93(1):23-9. Epub 2007 Mar 22. <u>Other child development tests used:</u> Griffith

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						modification. When creating developmental tools for a rural African setting, many items from Western tools can be adapted. The gross motor domain is more culturally adaptable, whereas social development is difficult to adapt and is culturally specific."			
Denver Developmental Materials	N/A	Middle East and North Africa (countries not specified in abstract)	Birth to six years	Not specified	Study to standardize the Denver Developmental Screening Test (DDST) on children from the Middle East and North Africa.	Age norms of developmental milestones on the personal-social, fine motor-adaptive, language, and gross motor skills are presented.	al-Naquib et al.	1999	al-Naquib, N., W. K. Frankenburg, et al. The standardization of the Denver Developmental Screening Test on Arab children from the Middle East and north Africa. <i>J Med Liban</i> . 1999;47(2): 95-106.
Denver Developmental Materials	N/A	Nepal	1 year	Not specified	Validation study of a scoring system previously validated in Cape Town used assess neonatal encephalopathy. A combination of the Denver Developmental Screening Test and Bailey 2 was used as a criterion measure.	Both schemes converted a pretest probability of 31% (the prevalence of major impairment at 1 y of age in this cohort) to a post-test probability of 55%. This showed only marginal improvement over the traditional risk marker of neurological abnormality at discharge (post-test probability 51%). There is difference both in practical or predictive terms whether one describes the neurological condition of the neonate using a descriptive or scoring system.	Ellis et al.	2001	Ellis, M, Shrestha L et al. Clinical predictors of outcome following mild and moderate neonatal encephalopathy in term newborns in Kathmandu, Nepal. <i>Acta Paediatr</i> . 2001; 90(3): 316-22.
Denver Developmental Materials	N/A	Singapore	3 months, 9 months, 17 months, 37 months and 60 months	Not specified	The Denver Developmental Screening Test, Singapore (DDST, Singapore), a Singapore version of the Denver II Developmental Screening Test, was field-tested with 2459 children.	Based on a single DDST, Singapore test result, there was a high false positive rate of 83.5% and lower false negative rate of 3.7% compared to the DAC assessment. However, the high false positive rate would be significantly reduced by doing a repeat screening test on the "questionable" cases and having the screened-out cases assessed by trained primary health care doctors. In conclusion, no major revision is needed in the scoring criteria suggested in the current version of DDST, Singapore. It is a useful tool in identifying children who will otherwise be missed without formal screening.	Lim et al.	1996	Lim HC, Ho LY, Goh LH, Ling SL, Heng R, Po GL. The field testing of Denver Developmental Screening Test Singapore: a Singapore version of Denver II Developmental Screening Test. <i>Ann Acad Med Singapore</i> . Mar 1996;25(2):200-209.
Denver Developmental Materials	N/A	Sweden	Not specified	Not specified	Study to investigate child self-rating of behavioural problems and competence as compared with their parents' ratings, in comparative samples of Kurdistanian refugee children in Sweden and a group of Swedish children.	When compared to the children's self-reported problems scores, the Swedish parents reported significantly lower scores than their children, compared to the Kurdistanian refugee parents.	Wahlsten et al.	2002	Wahlsten VS, Ahmad A, Von Knorring AL. Do Kurdistanian and Swedish parents and children differ in their rating of competence and behavioural problems? <i>Nord J Psychiatry</i> . 2002;56(4):279-283.
Denver Developmental Materials	N/A	Taiwan	Between ages 3 and 5	Chinese	A descriptive correlational study that used the Denver Developmental Screening	The DDST-Chinese version is valid in capturing the maturing status of this study's children. The fine-motor adaptive dimension was more advanced in	Chen CJ et al.	2003	Chen, C. J., I. C. Li, et al. Developmental status among 3 to 5-year-old preschool children

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					Test (DDST)-Chinese version to evaluate the child developmental status in three kindergartens in Peitou District of Taipei City.	this group than the DDST norms. There were no distinct differences between the study subjects and the language and person-social dimension norms. The differences between the study subjects and the fine motor-adaptive dimension norms suggest that the DDST-Chinese version norms need to be standardized in order to account for cultural differences. To be a valid tool, the DDST-Chinese version norms should be standardized to concur with current development among Taiwanese children.			in three kindergartens in the Peitou District of Taipei City. J Nurs Res. 2003; 11(2): 73-81.
Denver Developmental Materials	N/A	Taiwan	Between ages 3 and 6	Chinese	Study to evaluate and compare the growth and development of preschool children with congenital heart disease (CHD) to those of normal preschool children.	Preschoolers with CHD had more suspicious interpretations than non-CHD preschoolers, specifically in the language (P < 0.01) and gross motor sections (P < 0.001). Nevertheless, there were two items in the personal-social section and one in the language section on which the CHD children passed in the range of 55.6-63.2%. Problems were encountered with the Denver II test due to differences in language, culture and childrearing methods between Taiwanese and Western societies. These cultural differences must be considered when the test is used to assess development.	Chen CW et al.	2004	Chen CW, Li CY et al. Growth and development of children with congenital heart disease. J Adv Nurs. 2004; 47(3): 260-9.
Denver Developmental Materials	II	Taiwan	infants and children 0-5 (older than 5 included in the study as well)	Not specified in abstract or methods	To study "neurodevelopment and cognition in children after enterovirus 71 infection."	"Nine of the 16 patients with a poliomyelitis-like syndrome (56%) and 1 of the 5 patients with encephalomyelitis (20%) had sequelae involving limb weakness and atrophy. Eighteen of the 28 patients with cardiopulmonary failure after CNS involvement (64%) had limb weakness and atrophy, 17 (61%) required tube feeding, and 16 (57%) required ventilator support. Among patients who underwent DDST II assessment, delayed neurodevelopment was found in only 1 of 20 patients (5%) with severe CNS involvement and in 21 of 28 patients (75%) with cardiopulmonary failure (P<0.001 for the overall comparison). Children with cardiopulmonary failure after CNS involvement scored lower on intelligence tests than did children with CNS involvement alone (P=0.003). Enterovirus 71 infection with CNS involvement and cardiopulmonary failure may be associated with neurologic sequelae, delayed neurodevelopment, and reduced cognitive functioning. Children with CNS involvement without cardiopulmonary failure did well on neurodevelopment tests. (ClinicalTrials.gov number, NCT00172393 [ClinicalTrials.gov]). Copyright 2007 Massachusetts Medical Society."	Chang LY, Huang LM, Gau SS, Wu YY, Hsia SH, Fan TY, Lin KL, Huang YC, Lu CY, Lin TY.	2007	Chang LY, Huang LM, Gau SS, Wu YY, Hsia SH, Fan TY, Lin KL, Huang YC, Lu CY, Lin TY. Neurodevelopment and cognition in children after enterovirus 71 infection. N Engl J Med. 2007 Mar 22;356(12):1226-34. <u>Other child development tests used:</u> Wechsler Intelligence Scale
Denver	N/A	Turkey	?	Not specified	Study to analyze normative	Sex were similar to those found in other countries.	Epir et al.	1984	Epir S, Yalaz K. Urban Turkish

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Developmental Materials					data for 1176 healthy, urban Turkish children for sex and social-class differences in performance of the Denver Developmental Screening Test (DDST).	Social-class differences, in contrast to findings in other DDST studies, were consistently in favor of the more advantaged children in all areas of functioning from about the age of 10 months. These differences became more marked with age, particularly in language and fine-motor skills. These findings indicate considerable cultural differences within urban Turkey and suggest that the use of only one set of norms may lead to misdiagnosis. Further, they question the predictive validity of this test for lower-class urban Turkish children, particularly for language and fine-motor tasks.			children's performance on the Denver Developmental Screening Test. Dev Med Child Neurol. Oct 1984;26(5):632-643.
Denver Developmental Materials	N/A	Turkey	15 days to 72 months	Turkish	Study to find widely available, inexpensive, and non-invasive parameters for early identification or prediction of the infants with hypoxic-ischemic encephalopathy (HIE) who will have a severe adverse outcome (classified as death or a major neurological deficit).	DDST II at 6 months of age yielded a very high predictive accuracy (sensitivity=100%, specificity=95%). The DDST II at 6 months of age could be used in predicting severe neurological outcome in infants with HIE.	Hallioğlu et al.	2001	13. Hallioğlu O, Topaloglu AK, Zenciroğlu A, Duzovalı O, Yılgor E, Saribas S. Denver developmental screening test II for early identification of the infants who will develop major neurological deficit as a sequela of hypoxic-ischemic encephalopathy. Pediatr Int. Aug 2001;43(4):400-404.
Denver Developmental Materials	II	Turkey	age 40-75 months	Not specified in abstract and methods. Restandardized and revised for various populations	"In this study, children with an evidently recurrent otitis media were investigated. The present study examines the association between hearing loss versus developmental screening test parameters of preschool children."	"Language and verbal cognitive abilities were not affected significantly as a result of the presence of hearing loss because of OME. Using internationally standardized Denver-II test to evaluate the language development and other developmental screening parameters, no significant difference was found between the patient and control groups. This study failed to find any association between the hearing loss due to otitis media with effusion and speech and language parameters in preschool children."	Serbetcioglu B, Ugurtay O, Kirkim G, Mutlu B.	2008	Serbetcioglu B, Ugurtay O, Kirkim G, Mutlu B. No association between hearing loss due to bilateral otitis media with effusion and Denver-II test results in preschool children. Int J Pediatr Otorhinolaryngol. 2008 Feb;72(2):215-22. Epub 2007 Nov 28. <u>Other child development tests used:</u> N/A
Denver Developmental Materials	II	Turkey	assessed at 72-76 months old, then again 6-8 months later	Turkish translation and validation	To determine whether "preschool developmental screening" can "identify children at risk for school problems"	"All with normal/questionable results had satisfactory school performance while 26% of those with abnormal Denver II had low school scores. 43%, 27% and 0% of children with abnormal, questionable, and normal Denver II respectively had IQ<90 on WISC-R (p=0.05). Preschool Denver II can be recommended for populations at risk."	Bayoglu BU, Bakar EE, Kutlu M, Karabulut E, Anlar B.	2007	Bayoglu BU, Bakar EE, Kutlu M, Karabulut E, Anlar B. Can preschool developmental screening identify children at risk for school problems? Early Hum Dev. 2007 Sep;83(9):613-7. Epub 2007 Jan 25. <u>Other child development tests used:</u> N/A
Denver	II	Turkey	assessed at 3	Not specified	To investigate "the role of	"Seventeen stage I cases (47%), 12 stage II cases	Dağ Y, Fırat	2006	Dağ Y, Fırat AK, Karakaş HM,

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Developmental Materials			and 6 months old	in abstract or methods	DWI (Diffusion-weighted imaging) in the diagnosis and determining the prognosis of HIE (Hypoxic-ischemic encephalopathy) in neonates."	(33%), and 7 stage III cases (20%) were detected. DWI obtained within the first 24 hours showed high sensitivity (100%) in detecting the permanent neurological sequelae but with very low specificity (20%). The negative predictive value of DWI in this period was 100%; however, in DWI obtained at the end of the first month, not only its sensitivity was preserved, but its specificity reached 80%. The negative predictive value of DWI in this period was preserved and the positive predictive value improved. The importance of DWI in detecting sequelae at the end of the first month was also demonstrated by McNemar (p=0.250) and Kappa (Kappa=0.719) tests. There was no difference between conventional MR imaging and DWI in detecting sequelae at the end of first month. DWI is superior to other imaging modalities in detecting ischemia; not only because of its high sensitivity in the early phase, but also because of its high sensitivity and specificity in the late phase. Moreover, with its high negative predictive value, DWI can be used for excluding the possibility of sequelae development in the early phase of HIE cases for medico-legal purposes."	AK, Karakaş HM, Alkan A, Yakinci C, Erdem G.		Alkan A, Yakinci C, Erdem G. Clinical outcomes of neonatal hypoxic ischemic encephalopathy evaluated with diffusion-weighted magnetic resonance imaging. <i>Diagn Interv Radiol.</i> 2006 Sep;12(3):109-14. <u>Other child development tests used:</u> N/A
Denver Developmental Materials	N/A	United Arab Emirates	3 years	Arabic	Study examining the prevalence and psychosocial correlates of Developmental Language Delay (DLD) in the United Arab Emirates.	Of the 694 children screened for DLD, 69 (9.9%; CI 7.8-12.4) were found to have delays in the language sector of DDST. 45 (6.5%; CI 4.3-8.7) were identified as having general language disability, both in comprehension and expression. Language delay was found to be associated with rural living, mother being from a different nationality, non-involvement of domestic help in child care, family history of language delay, obstetric and perinatal problems and presence of behavioral problems in the child. Previous non-UAE nationality of the mother and total monthly family income emerged were related to general language delay.	Eapen et al.	2004	Eapen, V, Zoubeidi, T et al. Screening for language delay in the United Arab Emirates. <i>Child Care Health Dev.</i> 2004; 30(5): 541-9.
Denver Developmental Materials	N/A	Zaire	Under 2 years	Kikongo (dialect related to Kituba) or Lingala	Study to examine social and motor development deficits of fourteen asymptomatic HIV-infected Zairian children.	Fourteen asymptomatic HIV-infected Zairian children under 2 years of age displayed social and motor developmental deficits on the Denver Developmental Screening Test when compared with 20 HIV-negative cohorts born to HIV-infected mothers and 16 control children.	Boivin et al.	1995	Boivin, M. J., S. D. Green, et al. A preliminary evaluation of the cognitive and motor effects of pediatric HIV infection in Zairian children. <i>Health Psychol.</i> 1995; 14(1): 13-21
Denver Developmental Materials	Versions II and Developmental Screening Test		infants and children 0-5 (older than 5 included in the study as well)	Not specified in abstract (article in Russian)	To assess "psychomotor development in preschool and early preschool children, as a long term outcome of neonatal bacterial meningitis"	"We have determined that the results of neonatal bacterial meningitis are predictors of severe long-term outcome of the disease. Analysis of the obtained data enables us to conclude that outcome of neonatal bacterial meningitis and sepsis in combination with bacterial meningitis is more	Sirbiladze TsV, Tatishvili NA, Kipiani TB, Shvangiradze MSh, Kandareli LG.	2006	Sirbiladze TsV, Tatishvili NA, Kipiani TB, Shvangiradze MSh, Kandareli LG. [Assessment of psychomotor development in preschool and early preschool children, as a

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						severe than of bacterial sepsis."			long term outcome of neonatal bacterial meningitis] Georgian Med News. 2006 May;(134):90-3. <u>Other child development tests used:</u> Bayley, Wechsler Intelligence scale, Raven progressive matrices
Early Childhood Care and Development (ECCD) Checklist	N/A	Philippines	0 to 4 years	English and Pilipino	To evaluate an important ECD initiative of the Philippine government using longitudinal data collected that we collected over three years on a cohort of 6,693 children age 0-4 years at baseline in two "treatment" regions that received the ECD program and a "control" region that did not receive the intervention.	There is evidence of significant improvement in cognitive, social, and motor development scores for children age three and below who reside in ECD program areas compared to those in nonprogram areas. The proportion of children below age four with worms is significantly lower in program compared to non-program areas among those exposed to the program for the longest time interval (at least 17 months). Among children exposed to the program for at least 17 months, improvements in psycho-social development are evident across all age ranges.	UNICEF Philippines	2007	Armezin G., Behrman J., Duazo P., Ghuman S., Gultiano S., King E.M., Lee N. (2007). <i>Early childhood development programs and children's health, nutrition and psychosocial development</i> . Unpublished manuscript.
Early Development Inventory	N/A	Canada	4 to 5 years	English	Study to evaluate the psychometric properties of the EDI.	Inter-rater agreements moderate to high for teacher ratings, and low to moderate for parent ratings. Concurrent validity statistically significant, though ranged low to moderate.	Janus M & Offord DR	2007	Janus M & Offord DR. (2007). Development and psychometric properties of the Early Developmental Inventory (EDI): a measure of children's school readiness. <i>Canadian Journal of Behavioral Science</i> , 39(1), 1-22.
Escala Argentina de Inteligencia Sensorimotriz (EAIS)	N/A	Argentina	6 to 24 months	Spanish	To present a test which allows the behavior of our infants to be assessed within the psychomotor and cognitive development area. Continuous measures with Argentinian norms and cut-off points for developmental delay.	The Argentine Scale of Sensory Motor Intelligence (EAIS) was created for babies aged from 6 to 24 months old. Said scale was built considering 323 Argentine babies in order to know the level of mental development during the non-verbal period, obtain standard figures of reference of the Argentine population and contribute with a useful as well as effective tool to diagnose the intellectual development of Argentine babies aged from 6 to 24 months old and detect cognitive disorders.	Oiberman et al.	2006	Oiberman A, Orellana LC, and Mansilla M. (2006). Evaluacion de la inteligencia en bebes argentinos: Escala Argentina de Inteligencia Sensoriomotriz. <i>Arch Argentina Pediatrica</i> , 104(4), 316-324. OR Oiberman, A. (2005). Evaluación de la inteligencia en bebés argentinos: Escala Argentina de Inteligencia Sensorio-motriz. <i>Revista Argentina de Clínica Psicológica</i> , 14(3), 213-218.
Escala Argentina de Inteligencia	N/A	Argentina	0-24 months	Spanish	To present the results of the application of the Argentine Scale of Sensory Motor	11 out of 15 babies reached an adequate outcome for their ages in the EAIS, and in the EEDP their outcomes were also considered normal. Three	Oiberman	2006	Oiberman, A. (2006). Resiliencia y factores de protección en bebés

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Sensorimotriz (EAIS)					Intelligence (EAIS) and the Psychomotor Development Assessment (EEDP) in 15 babies, born before or during the confinement of mothers in prison in a penitentiary located in the province of Buenos Aires, Argentina.	infants show a psychomotor and intellectual development much lower than expected for their age, and 3 other babies reached a psychomotor and cognitive development much higher than the expected mean.			vulnerables. Aplicación de la Escala Argentina de Inteligencia Sensoriomotriz. Acta Psiquiátrica y Psicológica de America Latina, 52(1), 19-25.
Escala Argentina de Inteligencia Sensoriomotriz (EAIS)	N/A	Argentina	6 to 15 months	Spanish	To observe the influence of mother/child bonds on the sensorymotor intelligence in 6 to 15 month old babies.	Results show no significant differences in the mother/child bond over the baby's intelligence percentile, even though there is a slight tendency. There were high significant differences in months of gestation over the kind of bond, and a significant difference was also found in the mother's education level over the mother/child bond.	Rodriguez	2006	Rodríguez, G. (2006). Tipo de vínculo madre/hijo y desarrollo intelectual sensoriomotriz en niños de 6 a 15 meses de edad. <i>Interdisciplinaria Revista de Psicología y Ciencias Afines</i> , 23(2), 175-201.
Escala de Evaluación del Desarrollo Psicomotor (EEDP)	N/A	Argentina and Chile	0-2	Spanish	Determine whether the EEDP could be used to measure the quality of life for patients with mechanical ventilators	It was able to do this	Salinas et al	2008	Salinas P, Farias A, Gonzalez X, and Rodriguez C. (2008). Calidad de vida relacionada en salud: concepto y evaluación en pacientes con ventilación mecánica no invasiva. <i>Neumología Pediátrica</i> , 34-39.
Escala de Evaluación del Desarrollo Psicomotor	N/A	Argentina	0 to 5 years	Spanish	To determine age of attainment of developmental milestones of children of Argentina.	Comparative test showed no significant differences between tests. Multiple logistic regressions showed that social class, maternal education and sex (female) were associated with earlier attainment of some selected developmental items, achieved at ages later than 1 year. Selected items achieved before the first year of life were not affected by any of the independent environmental variables studied.	Lejarraga et al.	2002	Lejarraga H, Pascucci MC, Krupitzky S, Kelmansky D, Bianco A, Martinez E, Tibaldi F, Cameron N. Psychomotor development in Argentinean children aged 0-5 years. <i>Paediatr Perinat Epidemiol</i> . 2002 Jan;16(1):47-60.
Escala de Evaluación del Desarrollo Psicomotor (EEDP)	N/A	Argentina	6 to 24 months	Spanish	To present a test which allows the behavior of our infants to be assessed within the cognoscience development area.	The Argentine Scale of Sensory Motor Intelligence (EAIS) was created for babies aged from 6 to 24 months old. Said scale was built considering 323 Argentine babies in order to know the level of mental development during the non-verbal period, obtain standard figures of reference of the Argentine population and contribute with a useful as well as effective tool to diagnose the intellectual development of Argentine babies aged from 6 to 24 months old and detect cognitive disorders.	Oiberman et al.	2006	Oiberman A, Orellana LC, & Mansilla M. (2006). Evaluación de la inteligencia en bebés argentinos: Escala Argentina de Inteligencia Sensoriomotriz. <i>Arch Argentina Pediatría</i> , 104(4), 316-324. OR Oiberman, A. (2005). Evaluación de la inteligencia en bebés argentinos: Escala Argentina de Inteligencia Sensorio-motriz. <i>Revista Argentina de Clínica Psicológica</i> , 14(3), 213-218.
Escala de Evaluación del Desarrollo	N/A	Argentina	0-2	Spanish	Describe a series of tests of sensory motor intelligence in breastfeeding children	Defined a number of terms relevant to assessment of children 0-2, reviewed both English language and Spanish language tests and described their use.	Schapira	2007	Schapira IT. (2007). Comentarios y aportes sobre desarrollo e inteligencia

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Psicomotor (EEDP)						Mentioned EEDP and EAIS (see below).			sensoriomotriz en lactantes. Analisis de herramientas de evaluacion de uso frecuente. Actualizacion bibliografica. <i>Revista del Hospital Materno Infantil Ramon Sarda</i> , 26(1), 21-27.
Escala de Evaluacion del Desarrollo Psicomotor (EEDP)	N/A	Argentina	0-24 MONTHS	Spanish	To present the results of the application of the Argentine Scale of Sensory Motor Intelligence (EAIS) and the Psychomotor Development Assessment (EEDP) in 15 babies, born before or during the confinement of mothers in prison in a penitentiary located in the province of Buenos Aires, Argentina.	11 out of 15 babies reached an adequate outcome for their ages in the EAIS, and in the <i>EEDP</i> their outcomes were also considered normal. Three infants show a psychomotor and intellectual development much lower than expected for their age, and 3 other babies reached a psychomotor and cognitive development much higher than the expected mean.	Oiberman	2006	Oiberman, A. (2006). Resiliencia y factores de protección en bebés vulnerables. Aplicación de la Escala Argentina de Inteligencia Sensoriomotriz. <i>Acta Psiquiátrica y Psicológica de America Latina</i> , 52(1), 19-25.
Escala de Evaluacion del Desarrollo Psicomotor (EEDP)	N/A	Argentina and Chile		Spanish	Normed in Chile in 1976	Was able to be used as a measure of well-being for children with mechanical ventilation.	Salinas et al		Salinas P, Farias A, Gonzalez X, and Rodriguez C. (2008). Calidad de vida relacionada en salud: concepto y evaluacion en pacientes con ventilacion mecanica no invasiva. <i>Neumologia Pediatrica</i> , 34-39.
Guide for Monitoring Child Development	N/A	Turkey	0 to 3.5 years	Turkish and English	Study to report on the development and psychometric properties of the GMCD monitoring component for children aged 0 to 24 months.	Age-dependent attainment pattern was seen at all milestones. Interrater reliability and internal consistency were both high.	Ertem IO et al.	2008	Ertem IO, Dogan DG, Gok CG, Kizilates SU, Caliskan A, Atay G, Vatandas N, Karaaslan T, Baskan SG, & Ciccetti DV. (2008). A guide for monitoring child development in low- and middle- income countries. <i>Pediatrics</i> , 121, 581-589.
ICMR Psychosocial Developmental Screening Test	N/A	India	0 to 6 years	English and local language	Study to assess the psychosocial development of well nourished and malnourished children aged 0 to 6 years and to identify the microenvironmental factors influencing their growth and development.	Malnourished children attained milestones at a later age. Delays (to the extent of 7 to 11 months) especially observed in vision and fine motor, language and comprehension, and personal social. Factors contributing to positive psychosocial development include paternal involvement with childcare, both parents teaching child, small family size, and paternal occupation.	Vazir et al.	1998	Vazir S, Naidu AN, & Vidyasagar P. (1998) Nutritional status, psychosocial development and the home environment of Indian rural children. <i>Indian Pediatrics</i> , 35(10), 959-966.
ICMR Psychosocial Developmental Screening Test	N/A	India	0 to 6 years	English and local language	To develop a screening measure for the early detection of developmental disabilities in children under the age of 6.	The centile values for each milestone were similar to the WHO, a well established measure.	Vazir et al.	1992	Vazir S, Naidu AN, Vidyasagar P, Lansdown RG, & Reddy V. (1992). Screening test battery for assessment of psychological development. <i>Indian Pediatrics</i> , 31, 1465-1475.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
ICMR Psychosocial Developmental Screening Test	N/A	India	0 to 6 years	English and local language	The 50th centile values for each milestone of the ICMR were compared to those of three American and two Indian test batteries that are both used in India.	All of the scales show sufficient agreement between each other in each of the milestones.	Vazir et al.	1994	Vazir S, Lansdown R, Naidu AN, Vidyasagar P, & Reddy V. (1994). A comparison of Indian and American scales of child development. <i>Journal of the Indian Academy of Applied Psychology, 20(2)</i> , 175-181.
ICMR Psychosocial Developmental Screening Test	N/A	India	0 to 6 years	English and Telugu	To examine the informal preschool education provided by the government-sponsored Integrated Child Development Services (ICDS) for children residing in the 3 southern Indian states of Andhra Pradesh, Karnataka, and Tamil Nadu.	ICDS Subjects, even those who were malnourished, obtained motor and mental age scores comparable to their chronological ages, while non-attending control Ss lagged behind by several months. A significant percentage of Anganwadi Centre personnel displayed inattentive, uninformed, and inappropriate behavior.	Vazir et al.	1999	Vazir, S., & Kashinath, K. (1999). Influence of the ICDS on psychosocial development of rural children in Southern India. <i>Journal of the Indian Academy of Applied Psychology, 25(1)</i> , 11-24.
Infant and Toddler Socio-Emotional Assessment (ITSEA) (Chinese)	N/A	China	toddlers 12-36 months	Chinese version of Infant Toddler Social and Emotionall Assessment (CITSEA)	"To evaluate the reliability and validity of the Chinese version of Infant-Toddler Social and Emotional Assessment (CITSEA)."	"The reliability and validity of CITSEA were examined by standard psychometric methods. 1) The test-retest reliability of four broad domains ranged from 0.78 to 0.89 at the significant level $p < 0.001$; 2) Split-half reliability ranged from 0.82 to 0.90 ($p < 0.001$); 3) The alpha coefficient was noticed to range from 0.79 to 0.88, which demonstrated good internal consistency. Furthermore, as is hypothesized, the score of CITSEA domains was significantly correlated with subscale's score of the Child Behavior Checklist 2/3 (CBCL2/3) and dimension's score of China Toddler Temperament. Confirmatory factor analysis demonstrated a good and reliable match of the model, indicating that CITSEA outlines the social and emotional development of Chinese urban children aged 12 to 36 months. The Chinese version of ITSEA is valid and the psychometric properties of this translated version (including its reliability and validity) are at an acceptable standard. It can be used as an instrument for assessing social and emotional problems, including delays in social-emotional competence for Chinese young children."	Jianduan Z, Huishan W, Shuhua S, Xiaonan H, Guoyan L, Guangli L, Junxin S.	2009	Jianduan Z, Huishan W, Shuhua S, Xiaonan H, Guoyan L, Guangli L, Junxin S. Reliability and validity of standardized Chinese version of Urban Infant-Toddler Social and Emotional Assessment. <i>Early Hum Dev.</i> 2009 Jan 30. [Epub ahead of print] <u>Other child development tests used:</u> results correlated with CBCL 2/3
ITSEA	N/A	France	12-36 months	French	To document adaptation of the ITSEA (named the Evaluation Sociale et Emotionnelle de Jeunes Enfants (ESEJE)) for the French population.	All domains and most scales show adequate intrascale reliability. Certain age and gender effects on mean score differed from those reported for the U.S. sample. Our preliminary results indicate that the ESEJE shares the same latent factor structure as its English counterpart and may be a promising tool for the early detection of problem behaviors and delays in the acquisition of competencies in a French population.	Bracha et al.	2006	Bracha Z, Perez-Diaz F, Gerardin P, et al. A french adaptation of the infant-toddler social and emotional assessment. <i>Infant Mental Health Journal.</i> 2004;25:117-129.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
ITSEA	N/A	Romania	12-8 months	Romanian	Study to examine attachment in institutionalized and community children 12-31 months of age in Bucharest.	The ITSEA was administered to caregivers to examine a variety of problem behaviors and competencies of children. As predicted, children raised in institutions exhibited serious disturbances of attachment as assessed by all methods. Observed quality of caregiving was related to formation and organization of attachment in children living in institutions. These results held even when other variables, such as cognitive level, perceived competence, and quantitative interaction ratings, were controlled for.	Zeanah et al.	2005	Zeanah CH, Smyke AT, Koga SF, Carlson E. Attachment in institutionalized and community children in Romania. <i>Child Dev.</i> 2005;76:1015-1028.
Kaufman Assessment Battery for Children (KABC)	N/A	Benin	7-11 years	French	The aim of this study was to examine the effect of an improvement in iodine status on mental and psychomotor performance of schoolchildren (7–11 y) who were moderately to severely iodine deficient.	The study, which was originally planned as a double-blind, randomized, placebo-controlled intervention, was carried out in an iodine-deficient population of schoolchildren (n = 196) in northern Benin. As the population began to have access to iodized salt during the 1-y intervention period, the study population was split post hoc—on the basis of urinary iodine concentrations—into a group with improved iodine status and a group with unchanged iodine status. Changes in mental and psychomotor performance over the intervention period were compared. Children with increased urinary iodine concentrations had a significantly greater increase in performance on the combination of mental tests than did the group with no change in urinary iodine concentrations.	van den Briel et al.	2000	van den Briel T, West CE, Bleichrodt N, van de Vijver F, Ategbo EA, Hautvast JA. Improved iodine status is associated with improved mental performance of schoolchildren in Benin. <i>American Journal of Clinical Nutrition</i> , 2000, 72(5), 1179-1185.
Kaufman Assessment Battery for Children (KABC)	N/A	Germany	6.0 to 18.11 years	German	Study to analyze cognitive functions in a large group of patients with spinal muscular atrophy.	The standard score in the 'mental processing composite' scale of the Kaufman-ABC was identical in the spinal muscular atrophy group and controls (103.8). There were no significant differences in any of the tests among different grades of severity (spinal muscular atrophy types I-III). It can be concluded that children and adolescents with spinal muscular atrophy have a general intelligence in the normal range. By adolescence, environmentally mediated aspects of intelligence are higher in patients with spinal muscular atrophy. It could be speculated that the development of cognitive skills and knowledge is a creative way to compensate the many restrictions due to their physical handicap.	von Gontard et al.	2002	von Gontard A, Zerres K, Backes M, et al. Intelligence and cognitive function in children and adolescents with spinal muscular atrophy. <i>Neuromuscul Disord.</i> Feb 2002;12(2):130-136.
Kaufman Assessment Battery for Children (KABC)	N/A	Germany	6 years	German	Study to assess the prevalence of intellectual-, language-, and prereading-skill deficits in a geographically defined whole-population sample of very preterm children at 6 years of age in southern	There was a high prevalence of long-term multiple cognitive problems in very preterm children. These persistent cognitive problems appear to be of pre- or neonatal (treatment) rather than postnatal social origin.	Wolke et al.	1999	Wolke D, Meyer R. Cognitive status, language attainment, and prereading skills of 6-year-old very preterm children and their peers: the Bavarian Longitudinal Study. <i>Dev Med Child Neurol.</i> Feb 1999;41(2):94-109.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					Germany.				
Kaufman Assessment Battery for Children (KABC)	N/A	Laos & U.S.	Lao children:4.5 to 12.5 years; U.S. children: 5.5 to 12.5 years	Lao and English	Study to examine the relationships among Auditory-Verbal, Visual, and Cross-Modal Memory Span tasks with 46 Lao and 56 American children.	In younger Lao and American children, Auditory-Verbal Memory Span was not significantly correlated with Visual Memory Span or Cross-Modal Memory Span, whereas significant correlations were found between Auditory-Verbal Memory Span and these other tasks in older children from both countries. The similar pattern in both cultures, together with findings from Zaire, supports the hypothesis that the relationship between Auditory-Verbal and Visual Memory Span may change with reading training.	Conant et al.	2003	Conant LL, Fastenau PS, Giordani B, et al. Environmental influences on primary memory development: a cross-cultural study of memory span in Lao and American children. J Clin Exp Neuropsychol. Dec 2003;25(8):1102-1116.
Kaufman Assessment Battery for Children (KABC)	N/A	Netherlands	42 months	Dutch	Study to examine possible adverse effects of environmental exposure to polychlorinated biphenyls (PCB) and dioxins on cognitive functioning in young children.	Maternal PCB exposure was associated with lower scores on the overall cognitive and sequential and simultaneous processing scales of the Kaufman Assessment Battery for Children (all P < .05). The highest exposed group (SigmaPCB ≥ 3 microg/L) scored 4 points lower on all 3 scales of the K-ABC when compared with the lowest exposed group (SigmaPCB < 1.5 microg/L). Both lactational exposure and current exposure to PCBs and dioxins were not related to 42-month cognitive performance.	Patandin et al.	1993	Patandin S, Lanting CI, Mulder PG, Boersma ER, Sauer PJ, Weisglas-Kuperus N. Effects of environmental exposure to polychlorinated biphenyls and dioxins on cognitive abilities in Dutch children at 42 months of age. J Pediatr. Jan 1999;134(1):33-41.
Kaufman Assessment Battery for Children (KABC)	N/A	Senegal	5 to 12 years	Wolof	Study to assess the effects of early cerebral malaria on cognitive ability in Senegalese children.	Senegalese children with a history of cerebral malaria (CM) performed more poorly on the Kaufman Assessment Battery for Children (K-ABC) Simultaneous Processing domain and on the Test of Variables of Attention (TOVA) attention capacity indicators in comparison with a matched control group. CM can disrupt neuropsychological integration during critical developmental periods, impacting on global neurological integrity, attentional vigilance, perceptual acuity, and subsequent development of visual-spatial processing and memory foundational to global cognitive ability.	Boivin et al.	2002	Boivin MJ. Effects of early cerebral malaria on cognitive ability in Senegalese children. J Dev Behav Pediatr. Oct 2002;23(5):353-364.
Kaufman Assessment Battery for Children (KABC)	N/A	South Africa	Third grade children	Not specified	Study compared the performance of 21 black and 35 white third-grade South African children on the K-ABC and the Wechsler Intelligence Scale--Revised (WISC-R) at two schools for children with learning problems.	While the WISC-R Verbal and Full Scale IQ of black children were significantly lower than that of whites, there was no significant difference between these groups on the K-ABC. Support is provided for the usefulness of the K-ABC as a relatively nondiscriminatory alternative to the WISC-R for South African children.	Skuy M et al.[Also cited under WISC above]	2000	Skuy M, Taylor M, O'Carroll S, Fridjhon P, Rosenthal L. Performance of black and white South African children on the Wechsler Intelligence Scale for Children--Revised and the Kaufman Assessment Battery. Psychol Rep. 2000 Jun;86(3 Pt 1):727-37.
Kaufman Assessment Battery for	N/A	Zaire	primary school children (age not given in	Kikongo (dialect related to Kituba) or	Study to examine whether cognitive performance for schoolchildren in Zaire	With respect to our home evaluation for each child, factors related to the nutritional and economic well-being of the home environment proved a reliable	Boivin et al.	1993	Boivin MJ, Giordani B. Improvements in cognitive performance for schoolchildren

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Children (KABC)			abstract)	Lingala	improves following an iron supplement and treatment for intestinal parasites.	marker for Simultaneous Processing ability. However, the present findings also suggest that over the short-term, changes in blood Hgb that accompany both vermifuge and iron supplement treatment together can improve certain aspects of cognitive ability, perhaps by means of heightened attentional capacity.			in Zaire, Africa, following an iron supplement and treatment for intestinal parasites. <i>J Pediatr Psychol.</i> Apr 1993;18(2):249-264.
Kilifi	Inventory	Kenya	6 to 35 months	Kiswahili and local language	Study to evaluate the reliability, validity, and acceptability of the Kilifi Developmental Inventory (a modified version of the KDC), to evaluate the applicability in an urban community, to develop reference tables for the KDI and to assess its ability to identify children with developmental impairment.	The inventory showed excellent internal consistency, observer agreement, test-retest reliability, and sensitivity to maturational changes.	Abubakar et al.	2008	Abubakar A, Holding P, van Baar A, Newton CR, & van de Vijver FJ. (2008). Monitoring psychomotor development in a resource-limited setting: an evaluation of the Kilifi Developmental Inventory. <i>Annals of Tropical Paediatrics.</i> 28(3), 217-26.
Kilifi	Checklist	Kenya	7 months to 7 years	Kiswahili and local language	Study to develop a culturally-informed measure of developmental outcome and to apply it to detect differences in developmental level between children with cerebral malaria.	Measure demonstrated high internal consistency, good inter-observer reliability, age sensitivity and strong relations with parental report of child functioning.	Abubakar et al.	2007	Abubakar A, van de Vijver FJ, Mithwani S, Obiero E, Lewa N, Kenga S, Khamis K, & Holding P. (2007). Assessing developmental outcomes in children in Kilifi, Kenya, following prophylaxis for seizures in cerebral malaria. <i>Journal of Health Psychology.</i> 12(3), 417 - 430.
Kilifi	Inventory	Kenya	24 to 35 months	Kiswahili and local language	Study to determine if anthropometric status mediates the relation between socioeconomic status and psychomotor development of young children in resource-limited settings.	A significant association was found between anthropometric status (weight, height, and arm and head circumference) and psychomotor functioning, as well as between SES and anthropometric status. Results show that weight and height mediate relationship between psychomotor development and SES.	Abubakar et al.	2008	Abubakar A, Van de Vijver F, Van Baar A, Mbonani L, Kalu R, Newton C, & Holding P. (2008). Socioeconomic status, anthropometric status, and psychomotor development of Kenyan children from resource-limited settings: a path-analytic study. <i>Early Human Development.</i> 84(9), 613-21.
LeiterR or Leiter International Performance Scale	N/A	Canada	Not specified	French and English	Study to determine whether bilingual children with specific language impairment (SLI) are similar to monolingual age mates with SLI, in French and English.	The bilingual children with SLI in this study appeared similar to their monolingual peers for the aspects of grammatical morphology examined in each language. These bilingual-monolingual similarities point to the possibility that SLI may not be an impediment to learning two languages, at least in the domain of grammatical morphology.	Paradis et al.	2003	Paradis J, Crago M, Genesee F, Rice M. French-english bilingual children with SLI: How do they compare with their monolingual peers? <i>Journal of Speech, Language, and Hearing Research.</i> 2003;46:113-127.
LeiterR or	N/A	Italy	6 to 21 years	Italian	Study to investigate the	These data suggest that persons with congenital	Carlesimo et al.	2006	Carlesimo GA, Galloni F,

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Leiter International Performance Scale					nature of the articulatory rehearsal mechanism of the Articulatory Loop in Baddeley's Working Memory model. The participants' mental age was evaluated using the Leiter International Performance Scale.	anarthria/dysarthria and ID present defective maturation at many levels of the Articulatory Loop and reduced contribution of semantic-lexical processing in the temporary retention of phonological sequences. It is likely that in these individuals both communicative deficits and ID play a role in the impaired development of verbal short-term memory abilities.			Bonanni R, Sabbadini M. Verbal short-term memory in individuals with congenital articulatory disorders: New empirical data and review of the literature. <i>Journal of Intellectual Disability Research</i> . 2006;50:81-91.
LeiterR or Leiter International Performance Scale	N/A	Italy	3.6 to 6.6 years	Italian	Study to describe the clinical outcome of a consecutive sample of preschool children with pervasive developmental disorders (PDD) treated with risperidone monotherapy.	The study reports significant improvement at the last observation in Children's Psychiatric Rating Scale (CPRS) and Children's Global Assessment Scale (CGAS) scores.	Masi et al.	2003	Masi G, Cosenza A, Mucci M, Brovedani P. A 3-year naturalistic study of 53 preschool children with pervasive developmental disorders treated with risperidone. <i>J Clin Psychiatry</i> . 2003;64:1039-1047.
LeiterR or Leiter International Performance Scale	N/A	Italy	3 and 4 years	Italian	Study to investigate attention and perceptual and spatial working memory abilities in preterm, low birth weight preschool children without evident brain disorders as determined by normal cerebral ultrasound findings and normal motor development.	Children born prematurely without major neurological deficits and with a normal cognitive level may have specific difficulty in sustained attention, visuospatial processing, and spatial working memory when evaluated at ages 3-4.	Vicari et al.	2004	Vicari S, Caravale B, Carlesimo GA, Casadei AM, Allemand F. Spatial working memory deficits in children at ages 3-4 who were low birth weight, preterm infants. <i>Neuropsychology</i> . 2004;18:673-678.
LeiterR or Leiter International Performance Scale	N/A	Saudi Arabia	Less than 24 months	Not specified	Study to examine the cognitive delay of 44 Saudi Arabian children with unknown dysmorphic syndromes, and the relationship between such delay and particular patterns of dysmorphic features.	No significant relationship was found between various dysmorphic feature clusters and measured cognition.	Nester et al.	1992	Nester MJ, Sakati N, Greer W. Unknown dysmorphic syndromes and developmental delay in Saudi Arabia. <i>J Child Neurol</i> . 1992;7:S64-S68.
LeiterR or Leiter International Performance Scale	N/A	Spain	9 to 28 years	Spanish	Study to examine impairment displayed by adults with autism in nominating and identifying emotions.	Autistic subjects displayed significantly more difficulty in nominating and identifying emotions as compared to mentally retarded individuals and normal controls. Findings suggest that emotion nomination impairment is a specific and chronic deficit of autism.	García-Villamisar et al.	1999	García-Villamisar D, Polaino-Lorente A. Nomination and identification of emotions: A comparative study between persons with autism, mental retardation and general population/Nominación e identificación de emociones: Un estudio comparativo entre autistas, deficientes mentales y población general. <i>Estudios de Psicología</i> . 1999;20:33-44.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
LeiterR or Leiter International Performance Scale	N/A	Sweden	Pre-school age (exact age not specified)	Swedish and Arabic	Study to investigate grammatical development over 12 months in Swedish and Arabic in 10 Swedish-Arabic pre-school children with severe LI and 10 Swedish-Arabic pre-school children without LI.	Bilingual children, both with and without LI, developed grammatical structures in Swedish and Arabic in the same implicational way. Children with severe LI could develop two languages, although the pace of development was much slower in both languages. Bilingual children with severe LI were also more vulnerable to limited exposure of both their languages.	Salameh et al.	2004	Salameh E, Håkansson G, Nettelbladt U. Developmental perspectives on bilingual swedish-arabic children with and without language impairment: A longitudinal study. <i>International Journal of Language & Communication Disorders</i> . 2004;39:65-91.
LeiterR or Leiter International Performance Scale	N/A	Taiwan	3 years, 8 months and 3 years, 11 months at baseline.	Chinese	Longitudinal study to follow into childhood a cohort of preschool children, initially diagnosed as autistic or non-autistic with developmental delay, to evaluate the stability of cognitive assessments performed during the preschool period.	Mean developmental quotients (DQ)non-verbal IQ at initial assessment was 73.9±23.9 for the case group and 80.3±23.2 for the control group. Intellectual functioning can be a valid measure in Taiwanese preschool children with autism, and has an equivalent meaning for children with autism and for non-autistic children with developmental delay.	Yang et al.	2003	Yang P, Jong Y, Hsu H, Chen C. Preschool children with autism spectrum disorders in taiwan: Follow-up of cognitive assessment to early school age. <i>Brain Dev</i> . 2003;25:549-554.
MacArthur Child Development Inventory (CDI)	N/A	China	children, age not specified in abstract	Mandarin. No published CDI norms for Mandarin Chinese at time of publication; this study developed a Mandarin early vocabulary inventory	"In this article, we report our efforts in developing this instrument, and discuss the data collected from 884 Chinese families in Beijing over a period of 12-30 months, based on our instrument."	"Chinese children's receptive and expressive lexicons as assessed by our inventory match well with those reported for English on the basis of CDI. In particular, our data indicate comprehension-production differences, individual differences in early comprehension and in later production, and different lexical development profiles among infants versus toddlers. We also make the checklists and norms of our inventory available to the research community via the Internet; they may be accessed from the Psychonomic Society's Archive of Norms, Stimuli, and Data, at www.psychonomic.org/archive ."	Hao M, Shu H, Xing A, Li P.	2008	Hao M, Shu H, Xing A, Li P. Early vocabulary inventory for Mandarin Chinese. <i>Behav Res Methods</i> . 2008 Aug;40(3):728-33. <u>Other child development tests used:</u> N/A
MacArthur Child Development Inventory (CDI)	N/A	Italy and UK	Between 1.6 and 2.6 years	Italian and English	The relationship between grammatical and lexical development was compared in 233 English and 233 Italian children, matched for age, gender, and vocabulary size on the MacArthur Communicative Development Inventories (CDI).	Italians had longer Mean Length of Utterances (MLUs) on most measures, but the ratio of actual to target MLUs did not differ between languages. Age and vocabulary both contributed significant variance to MLU, but the contribution of vocabulary was much larger, suggesting that vocabulary size may provide a better basis for crosslinguistic comparisons of grammatical development. The relationship between MLU and vocabulary size was non-linear in English but linear in Italian, suggesting that grammar 'gets off the ground' earlier in a richly inflected language. A possible mechanism to account for this difference is discussed.	Devescovi et al.	2005	Devescovi A, Caselli MC, Marchione D, Pasqualetti P, Reilly J, Bates E. A crosslinguistic study of the relationship between grammar and lexical development. <i>J Child Lang</i> . Nov 2005;32(4):759-786.
MacArthur Child Development	N/A	Netherlands	1 to 1.7 years	Dutch	Study to examine inter- and intra-family variation of MCDI scores when more	The greater the child's communicative ability, as rated by any one reporter, the more differences tended to emerge between reporters. In order to take	De Houwer et al.	2005	De Houwer A, Bornstein MH, Leach DB. Assessing early communicative ability: a cross-

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Inventory (CDI)					than one reporter completes the Infant Form 'Words and Gestures' of the Dutch adaptation of the MacArthur Communicative Development Inventory for the same child.	into account multiple reporters' assessments of the same child, we propose the use of a Cumulative CDI Score that credits the child with the best score for any item on the CDI as checked by any single reporter. We conclude that single reporter CDI reports may underestimate the child's communicative knowledge.			reporter cumulative score for the MacArthur CDI. J Child Lang. Nov 2005;32(4):735-758.
MacArthur Child Development Inventory (CDI)	N/A	Netherlands	17 months	Dutch	An ongoing longitudinal dyslexia research project in which vocabulary composition is investigated in 17-month-old children.	The Dutch children show the same general acquisition pattern as documented for other languages, but there are significant differences between the two groups of 17-month-old children (a group of children who are genetically at risk for dyslexia and a control group) in total number of words produced and in the linguistic composition of their productive vocabulary.	Koster et al.	2005	Koster C, Been PH, Krikhaar EM, Zwarts F, Diepstra HD, Van Leeuwen TH. Differences at 17 months: productive language patterns in infants at familial risk for dyslexia and typically developing infants. J Speech Lang Hear Res. Apr 2005;48(2):426-438.
MacArthur Child Development Inventory (CDI)	N/A	New Zealand	1.7; 2.1; 2.8; 3.4 years	New Zealand English	Study to assess the long-term predictive validity of the MacArthur Communicative Development Inventories: Words and Sentences (CDI:WS) for children's expressive and receptive vocabulary development.	Excellent reliability and good predictive validity was obtained for the NZ CDI:WS even over a 21-month delay. Predictive validity of the NZ CDI:WS for the PPVT-III was higher for children of mothers with less education.	Reese et al.	2000	Reese E, Read S. Predictive validity of the New Zealand MacArthur Communicative Development Inventory: Words and Sentences. J Child Lang. 2000 Jun;27(2):255-66.
MacArthur Child Development Inventory (CDI)	N/A	Sweden	18 months and 3 years	Swedish	Study to evaluate the effectiveness of a screening instrument (the Swedish Communication Screening at 18 months of age; SCS18), derived from the Swedish MacArthur-Bates Communicative Development Inventory, in identification of 18-month-old children who will be severely language disabled by 3 years of age.	Number of spoken words yielded the best prediction, and SCS18 was superior to the traditional method. A sensitivity of 50%, however, was not enough, and a stricter criterion resulted in too many false positives to be acceptable as routine. CONCLUSION: Although the SCS18 has strength, the age of 18 months seems to be too early for identification of severe language disability.	Westerlund et al.	2006	Westerlund M, Berglund E, Eriksson M. Can severely language delayed 3-year-olds be identified at 18 months? Evaluation of a screening version of the MacArthur-Bates Communicative Development Inventories. J Speech Lang Hear Res. Apr 2006;49(2):237-247.
McCarthy Scales of Children's Abilities (MSCA)	N/A	5 European countries: Belgium, Denmark, Greece, Sweden, and the United Kingdom	5 years	English, Dutch, Danish, Swedish, or Greek	Study to shed more light on the cognitive and motor development of 5-year-old intracytoplasmic sperm injection (ICSI-conceived) children.	The motor and cognitive development of ICSI-conceived children is very similar to that of naturally conceived children. However, demographic factors such as maternal educational level and maternal age at the time of the birth might play different roles in the cognitive development of ICSI and in vitro fertilization (IVF)-conceived children, compared with NC children.	Ponjaert-Kristoffersen et al.	2005	Ponjaert-Kristoffersen I, Bonduelle M, Barnes J, et al. International collaborative study of intracytoplasmic sperm injection-conceived, in vitro fertilization-conceived, and naturally conceived 5-year-old child outcomes: cognitive and motor assessments. Pediatrics. Mar 2005;115(3):e283-289.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
McCarthy Scales of Children's Abilities (MSCA)	N/A	Australia	4 years	Not specified	Study to assess the relationship of birthweight to cognitive development	The mean birthweight was 3386 g (SD: 517). There was a statistically significant association between birthweight and cognitive performance at age 2 years (adjusted deficit: 0.97 points per 100 g lighter; 95% CI: 0.4-1.5), but the magnitude of this association gradually decreased and became statistically non-significant at later childhood. The relationship between birthweight and cognitive development becomes progressively attenuated at increasing age. At older ages, socioenvironmental factors appear to play an increasingly important part in children's cognitive development.	Tong et al.	2006	Tong S, Baghurst P, McMichael A. Birthweight and cognitive development during childhood. <i>J Paediatr Child Health.</i> Mar 2006;42(3):98-103.
McCarthy Scales of Children's Abilities (MSCA)	N/A	France	4.5 years	French	Longitudinal study to investigate the effect of moderate prenatal alcohol exposure on psychomotor development of preschool-age children.	Consumption of 1.5 oz of absolute alcohol (approximately three drinks) or more per day during the first trimester of pregnancy was significantly related to a decrease of 7 points in the mean score on the general cognitive index of the McCarthy scales, after gender, birth order, maternal education, score for family stimulation, family status, maternal employment, child's age at examination, and examiner were controlled for. This study showed that moderate to heavy alcohol consumption during pregnancy, at levels well below those associated with fetal alcohol syndrome, has effects on children's psychomotor development.	Larroque et al.	1995	Larroque B, Kaminski M, Dehaene P, Subtil D, Delfosse MJ, Querleu D. Moderate prenatal alcohol exposure and psychomotor development at preschool age. <i>Am J Public Health.</i> Dec 1995;85(12):1654-1661.
McCarthy Scales of Children's Abilities (MSCA)	N/A	France	7 years	French	Study to assess the relationship between cognitive and behavioral assessments at age 7 years and the schooling, cognitive performance, and manual dexterity at age 17 years.	After controlling for effects of parental education and IQ, a negative teachers' rating of children's behaviour and abilities in first-grade (7 years) was correlated with early cessation of schooling, but also, unexpectedly, with higher scores for manual dexterity at 17 years. Manual dexterity was not related to cognitive performance at 17 years. It is suggested that the relationship between manual and cognitive performance varies during development. Although manual exploratory behaviour is an important correlate of early cognitive development, manual dexterity is probably not related to later academic	Takser et al.	2002	Takser L, Dellatolas G, Bowler R, Laplante N, Huel G. Predictive factors of manual dexterity and cognitive performance at 17 years: a 10-year longitudinal study in a rural area of France. <i>Percept Mot Skills.</i> Aug 2002;95(1):15-26.
McCarthy Scales of Children's Abilities (MSCA)	N/A	Jamaica	2.5 years to 8.5 years	Not specified	Study to evaluate the effects of cannabis consumption during pregnancy and lactation on infants from birth to school age in rural communities.	Most scores fell in the middle range of about 4, similar to the North American scores, except for the lower mean in the category of Threshold of Responsiveness, because of an unanticipated cultural difference. The adjustments made did not compromise the comparability of the findings.	Dreher et al.	1993	Dreher MC, Hayes JS. Triangulation in cross-cultural research of child development in Jamaica. <i>West J Nurs Res.</i> 1993 Apr;15(2):216-29.
McCarthy Scales of Children's Abilities	N/A	Jamaica	4 and 5 years	Not specified; English?	Study to evaluate the cognitive development of 59 Jamaican children, from birth to age 5 years, whose	No significant differences in developmental testing outcomes between children of marijuana-using and non-using mothers except at 30 days of age when the babies of users had more favourable scores on	Hayes JS et al.	1991	Hayes JS, Lampart R, Dreher MC, Morgan L. Five-year follow-up of rural Jamaican children whose mothers used

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
(MSCA)					mothers used marijuana during pregnancy.	two clusters of the Brazelton Scales: autonomic stability and reflexes. The developmental scores at ages 4 and 5 years were significantly correlated to certain aspects of the home environment and to regularity of basic school (preschool) attendance.			marijuana during pregnancy. West Indian Med J. Sep 1991;40(3):120-123.
McCarthy Scales of Children's Abilities (MSCA)	N/A	Mexico	children 36 months - 5 years (older than 5 included in the study as well)	Not specified for MSCA but Spanish language versions were used for the other development tests in the study.	"Our aim was to disaggregate the effects of cash transfer from those of other programme components."	"A doubling of cash transfers was associated with higher height-for-age Z score (beta 0.20, 95% CI 0.09-0.30; p<0.0001), lower prevalence of stunting (-0.10, -0.16 to -0.05; p<0.0001), lower body-mass index for age percentile (-2.85, -5.54 to -0.15; p=0.04), and lower prevalence of being overweight (-0.08, -0.13 to -0.03; p=0.001). A doubling of cash transfers was also associated with children doing better on a scale of motor development, three scales of cognitive development, and with receptive language. Our results suggest that the cash transfer component of Oportunidades is associated with better outcomes in child health, growth, and development."	Fernald LC, Gertler PJ, Neufeld LM.	2008	Fernald LC, Gertler PJ, Neufeld LM. Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of Mexico's Oportunidades. Lancet. 2008 Mar 8;371(9615):828-37. <u>Other child development tests used:</u> Woodcock Munoz -Spanish, Peabody Picture Vocabulary Test validated Spanish language version
McCarthy Scales of Children's Abilities (MSCA)	N/A	Seychelles Islands	66 months	Creole (French-based)	Study of the association between methylmercury (MeHg) exposure and the developmental outcomes of children in the Republic of Seychelles at 66 months of age.	No adverse outcomes at 66 months were associated with either prenatal or postnatal MeHg exposure. CONCLUSION: In the population studied, consumption of a diet high in ocean fish appears to pose no threat to developmental outcomes through 66 months of age.	Davidson et al.	1998	Davidson PW, Myers GJ, Cox C, et al. Effects of prenatal and postnatal methylmercury exposure from fish consumption on neurodevelopment: outcomes at 66 months of age in the Seychelles Child Development Study. Jama. Aug 26 1998;280(8):701-707.
McCarthy Scales of Children's Abilities (MSCA)	N/A	Seychelles Islands	66 months	Creole (French-based)	Study to examine association between prenatal or postnatal methylmercury exposure and child development.	A consistent pattern of effect modification (EM) of social and environmental factors has not been observed, suggesting that the results may be due to chance.	Davidson et al.	2004	Davidson PW, Myers GJ, Shamlaye C, Cox C, Wilding GE. Prenatal exposure to methylmercury and child development: influence of social factors. Neurotoxicol Teratol. Jul-Aug 2004;26(4):553-559.
McCarthy Scales of Children's Abilities (MSCA)	N/A	Seychelles Islands	66 months	Creole (French-based)	In this longitudinal study of prenatal and postnatal MeHg exposure from fish consumption and development, the McCarthy Scales of Children's Abilities (MSCA) were administered to children to assess relationship between prenatal exposure to	Analyses of both the standard and the recombined MSCA subscales showed no adverse associations with MeHg exposure and neuropsychological endpoints. A positive association between postnatal MeHg exposure and performance on the MSCA Memory subscale was found. These findings are consistent with previous reports from the Seychelles study in that no adverse effects of MeHg exposure from fish consumption can be detected in this cohort.	Palumbo et al.	2000	Palumbo DR, Cox C, Davidson PW, et al. Association between prenatal exposure to methylmercury and cognitive functioning in Seychellois children: a reanalysis of the McCarthy Scales of Children's Ability from the main cohort study. Environ Res. Oct 2000;84(2):81-88.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					methylmercury and cognitive functioning.				
Milestones Checklists (Parent Evaluation of Developmental Status)	N/A	Australia	Kindergarten-aged children	English	Evaluated children's development (some with parent-reported developmental concerns) at the start of kindergarten using parent-report checklists to determine whether parent-report across multiple domains of development would predict school functioning.	Children whose parents reported concerns about development had somewhat lower scores two years later, but the sensitivity and specificity do not suggest that this measure should be used alone	Wake, Gerner, Gallagher	2005	Wake M, Gerner B, Gallagher S. Does parents' evaluation of developmental status at school entry predict language, achievement, and quality of life 2 years later? <i>Ambulatory Pediatrics</i> , 2005, 5(3), 143-149.
Multiple assessments (Ravens, modified Stroop, pegboard)		Bangladesh	Children in grades 1 and 2	Bangladeshi	Comparison of children with low thyroxine with those without.	Children with hypothyroidism performed less well on reading, spelling and general cognitive factor (factoring of Stroop, pegboard) than children without.	Huda et al.	1999	Huda S, Grantham-Macgregor S., Rahman K, Tompkins A. Biochemical Hypothyroidism Secondary to Iodine Deficiency Is Associated with Poor School Achievement and Cognition in Bangladeshi Children. <i>Journal of Nutrition</i> . 1999;129:980-987
Parental Report Scale	N/A	Tanzania	6 to 59 months	English and other	To measure the effects of iron supplementation and anthelmintic treatment on iron status, anaemia, growth, morbidity, and development of children aged 6 to 59 months.	Iron supplementation improved language development by 0.8 points (on 20 pt scale) as well as motor development by 1.1 points (on 18 pt scale).	Stoltzfus et al.	2001	Stoltzfus RJ, Kvalsvig JD, Chwaya HM, Montresor A, Albonico M, Tielsch JM, Saviolo L, and Pollitt E. (2001). Effects of iron supplementation and anthelmintic treatment on motor and language development of preschool children in Zanzibar: double blind, placebo controlled study. <i>BMJ</i> , 323, 1-8.
Peabody Picture Vocabulary Test (PPVT)	N/A	Brazil	children 4-5 years old (older than 5 included in the study as well)	Not specified in abstract	"To verify the performance of children with D-CP (spastic diplegic cerebral palsy) regarding their psycholinguistic skills."	"The results pointed to a significant correlation between the cognitive performance and the PPVT. The same correlation was confirmed between the PPVT and the subtest of auditory reception of the ITPA, when considering the psycholinguistic age. In the comparison between the auditory and visual abilities, the participants demonstrated a better performance in the activities that involved visual abilities, indicating a significant correlation in the association subtest. There was a statistically significant correlation between the level of motor impairment and the psycholinguistic performance, confirming the influence of the motor impairment in these activities. The children with D-CP and with suggestive signs of PVL (periventricular leukomalacia) presented deficits in the	Lamônica DA, Ferraz PM.	2007	Lamônica DA, Ferraz PM. [Periventricular leukomalacia and spastic diplegia: implications in the psycholinguistic abilities] <i>Pro Fono</i> . 2007 Oct-Dec;19(4):357-62. <u>Other child development tests used:</u> Illinois Test of Psycholinguistic Abilities (ITPA)

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						psycholinguistics abilities, justifying the need of additional studies in this area in order to investigate the development of these abilities."			
Peabody Picture Vocabulary Test (PPVT)	N/A	Chile	3-5 years; also 7th grade children (average age 13.61 years and 13.69 years)	Spanish	Used test whether rural children start school with significant handicaps in the acquirement of the spanish vocabulary when compared with their urban peers.	Average scores for the PPVT among rural children at lower socio- economic levels (age 3-5 years) were lower than among urban children at higher socio-economic levels.	Munoz et al.	1989	Munoz F, Quilodran C, Velasquez P, Niedmann C, Baeza A, Silva G, Osorio M. Acquirement of the Spanish vocabulary among rural and urban students of the 9th region. Rev Chil Pediatr. 1989 Nov-Dec;60(6):354-8. Spanish.
Peabody Picture Vocabulary Test (PPVT)	N/A	China	children 2-5 years old (older than 5 included in the study as well)	Not specified in abstract	"To explore the characteristics of visual search attention in children with Williams syndrome (WS)."	"(1) In the single-target searching task with no target similar distractor, the accurate response rate of the WS patients was 71% +/- 25%, significantly lower than that of the CA-matched children (87% +/- 14%, P = 0.001). The searching time of the WS children was 5 s, significantly longer than that of the CA-matched children (3 s, P = 0.000). The distance for touch of the WS children was 25 relative units, significantly longer than that of the CA-matched children (18 relative units, P = 0.000). The mean reaction time of the WS children was 5 s, significantly longer than that of the children with Down's syndrome (3 s, P = 0.022). The shape error rate of the WS children was 15%, significantly higher than that of the CA-matched children (0, P = 0.000). When non-targets similar to the targets were added, the accurate response rate of the WS group was 39% +/- 20%, significantly lower than those of the CA-matched and MA-matched children (77% +/- 23% and 66% +/- 23% respectively, both P = 0.000); the mean searching time of the WS children was 13 s, significantly longer than those of the CA-matched and MA-matched children (3 s and 5 s respectively, both P = 0.000); and the distance per touch of the WS children was 41 relative units, significantly longer than those of the CA-matched and MA-matched children (20 and 27 relative units, P = 0.000 and P = 0.004).	Xie CH, Shao J, Qin YF, Yang JB, Wang YX, Li R, Zhao ZY.	2008	Xie CH, Shao J, Qin YF, Yang JB, Wang YX, Li R, Zhao ZY. Visual search attention in children with Williams syndrome]. Zhonghua Yi Xue Za Zhi. 2008 Mar 11;88(10):679-83. <u>Other child development tests used:</u> N/A
						(2) The results of the dual target tasks showed that the accurate response rate of the WS children was 52% +/- 28%, significantly lower than that of the CA-matched children (78% +/- 22%, P = 0.000), the mean searching time of the WS children was 11 s, significantly longer than that of the CA-matched children (4 s, P = 0.000); and the distance per touch of the WS children was 54 +/- 27 relative units, significantly longer than that of the CA-matched			

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						children (31 +/- 13 relative units, P = 0.000). However, there were not significant differences in the accurate response rate, mean searching time, and distance per touch between the WS and DS children. The switching error rate of the WS children was 13%, significantly higher than those of the CA and MA-matched children (0 and 4% respectively, P = 0.000 and P = 0.004). (3) The vigilance task test showed that the accurate response rate of the WS children was 52% +/- 25%, significantly lower than that of the CA-matched children (80% +/- 21%, P = 0.000); the mean searching time of the WS children was 4 s +/- 1 s, significantly longer than that of the CA-matched children (3 s +/- 1 s, P = 0.000); and the error hit number of the WS children was 8, significantly more than that of the CA-matched children (3, P = 0.000). However, there were not significant differences in the accurate response rate, mean searching time, and error hit number between the WS and DS children. Distinct visual search deficits exist in WS children."			
Peabody Picture Vocabulary Test (PPVT)	N/A	China	children, age not quantified in abstract	Not specified in abstract	"To find out an effective therapy for autism."	"The total effective rate was 86.7% in the EA plus behavior therapy group which was better than 56.7% of the behavior therapy group, and had significant enhancement in sensation, association, body, and ability of self-care (P < 0.05) and was better than the behavior therapy group in sensation, body and self-care factors, with no significantly improvement in the scores of PPVT in the two groups (P > 0.05). EA (electroacupuncture) combined with behavior therapy can significantly improve clinical symptoms of autism, but does not improve intelligence."	Wang CN, Liu Y, Wei XH, Li LX.	2007	Wang CN, Liu Y, Wei XH, Li LX. [Effects of electroacupuncture combined with behavior therapy on intelligence and behavior of children of autism] Zhongguo Zhen Jiu. 2007 Sep;27(9):660-2. <u>Other child development tests used:</u> behavior ability tested but test not specified in abstract
Peabody Picture Vocabulary Test (PPVT)	N/A	China	children 20-60 months (older than 5 included in the study as well)	Not specified in abstract	"To explore the social adjustment status and affected factors thereof in Down syndrome children."	"There were no differences between the Down syndrome children and mental age-matched group in communication and socialization. The Down syndrome children were better than the mental age-matched group in self-help [(20.0 +/- 4.8) vs (13.3 +/- 4.7), t = 5.72, P = 0.000]; locomotion [(7.5 +/- 1.4) vs (6.4 +/- 1.6), t = 3.10, P = 0.003]; occupation [(8.2 +/- 2.4) vs (6.2 +/- 2.0), t = 3.68, P = 0.000], and self-direction [(5.9 +/- 2.6) vs (4.6 +/- 2.0), t = 2.28, P = 0.026]. The chronological age-matched group were much better than the Down syndrome children in all factors of social-life abilities, including self-help [(20.0 +/- 4.8) vs (26.5 +/- 4.9), t = 5.84, P = 0.000]; locomotion [(7.5 +/-	Wang YX, Li R, Shao J, Mao SS, Xie CH, Zhan JY, Qin YF, Zhu ZW, Zhao ZY.	2007	Wang YX, Li R, Shao J, Mao SS, Xie CH, Zhan JY, Qin YF, Zhu ZW, Zhao ZY. [A study on the social adjustment and its affective factors in Down syndrome children] Zhonghua Yi Xue Za Zhi. 2007 May 29;87(20):1402-6. <u>Other child development tests used:</u> Infants-Junior Middle School Students' Social-Life Abilities Scale

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						<p>1.4) vs (11.4 +/- 3.3), t = 6.76, P = 0.000]; occupation [(8.2 +/- 2.4) vs (14.4 +/- 3.9), t = 8.55, P = 0.000]; communication [(8.3 +/- 3.6) vs (18.3 +/- 4.8), t = 10.38, P = 0.000]; socialization [(9.6 +/- 2.3) vs (17.1 +/- 4.2), t = 9.76, P = 0.000], and self-direction [(5.9 +/- 2.6) vs (13.8 +/- 4.6), t = 9.25, P = 0.000]. There was a relationship between the raw score of PPVT and social-life abilities in these children.</p> <p>Even after controlling effect of age, it was still associated with self-help (r = 0.70, P = 0.000), locomotion (r = 0.74, P = 0.000), occupation (r = 0.77, P = 0.000), communication (r = 0.86, P = 0.000), socialization (r = 0.80, P = 0.000), and self-direction (r = 0.76, P = 0.000). Multiple stepwise regression showed that the main factor influencing self-help was family structure. Family structure and mother's education influenced locomotion. Family structure and newborn history influenced occupation, communication, socialization and self-direction. Down syndrome children have better social adjustment than the mental age-matched group, yet worse than the chronological age-matched group. Cognition development, family environment and newborn history differently influence the Down syndrome child, which means proper intervention can improve their social adjustment."</p>			
Peabody Picture Vocabulary Test (PPVT)	N/A	China	children 20-60 months (older than 5 included in the study as well)	Chinese revised version of the PPVT	"To evaluate social adjustment and related factors among Chinese children with Down syndrome (DS)."	<p>"There was no difference between the DS group and the MA group in terms of communication skills. However, the DS group scored much better than the MA group in self-dependence, locomotion, work skills, socialization and self-management. Children in the CA group achieved significantly higher scores in all aspects of social adjustment than the DS children. Partial correlations indicate a relationship between social adjustment and the PPVT raw score and also between social adjustment and age (significant r ranging between 0.24 and 0.92). A stepwise linear regression analysis showed that family structure was the main predictor of social adjustment. Newborn history was also a predictor of work skills, communication, socialization and self-management. Parental education was found to account for 8% of self-dependence. Maternal education explained 6% of the variation in locomotion. Although limited by the small sample size, these results indicate that Chinese DS children have better social adjustment skills when compared to their mental-age-matched</p>	Wang YX, Mao SS, Xie CH, Qin YF, Zhu ZW, Zhan JY, Shao J, Li R, Zhao ZY.	2007	<p>Wang YX, Mao SS, Xie CH, Qin YF, Zhu ZW, Zhan JY, Shao J, Li R, Zhao ZY. Study on the social adaptation of Chinese children with down syndrome. <i>Yonsei Med J.</i> 2007 Jun 30;48(3):412-20.</p> <p><u>Other child development tests used:</u> Infants-Junior Middle School Students' Social-Life Abilities Scale</p>

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						normally-developing peers, but that the Chinese DS children showed aspects of adaptive development that differed from Western DS children. Analyses of factors related to social adjustment suggest that effective early intervention may improve social adaptability."			
Peabody Picture Vocabulary Test (PPVT)	N/A	France	Children and adolescents (age not specified in abstract)	French	Used to examine the association of chronological age with syntax and vocabulary comprehension in children and adolescents with mental retardation.	Chronological age-related experience is linked to on the receptive vocabulary of children and adolescents with mental retardation.	Facon B et al.	2002	Facon B, Facon-Bollengier T, Grubar JC. Chronological age, receptive vocabulary, and syntax comprehension in children and adolescents with mental retardation. <i>Am J Ment Retard.</i> 2002 Mar;107(2):91-8.
Peabody Picture Vocabulary Test (PPVT)	N/A	New Zealand	1.7; 2.1; 2.8; 3.4 years	New Zealand English	Used as criterion measure to test predictive validity of NZ CDI:WS (New Zealand Version of MacArthur CDI).	New Zealand: Predictive validity of the NZ CDI:WS for the PPVT-III was higher for children of mothers with less education.	Reese et al.	2000	Reese E, Read S. Predictive validity of the New Zealand MacArthur Communicative Development Inventory: Words and Sentences. <i>J Child Lang.</i> 2000 Jun;27(2):255-66.
Peabody Picture Vocabulary Test (PPVT)	N/A	South Africa	Not specified	North-Sotho	Article describes process of culturally and linguistically adapting this assessment.	The findings of the research include practical examples of methodological considerations which should be taken into account while translating and undertaking cultural adaptation of test material. The newly adapted test material was also applied to a sample of 152 North-Sotho speaking pupils in the study area and the test results are discussed.	Pakendorf et al.	1997	Pakendorf C, Alant E. Culturally valid assessment tools: northern Sotho translation of the Peabody Picture Vocabulary Test-Revised. <i>S Afr J Commun Disord.</i> 1997;44:3-12
Peabody Picture Vocabulary Test (PPVT)	N/A	West Indies	Between 6 & 24 mos.; then periodically up to 14 years	Not specified	Assessing developmental levels of severely malnourished children participating in 3-year home visiting program.	Malnourished children participating in 3-year year home-visiting program (a psychosocial intervention) performed better on assessment than control group of malnourished children but not as well as adequately nourished controls.	Grantham-McGregor et al.	1994	Grantham-McGregor S, Powell C, Walker S, Chang S, Fletcher P. The long-term follow-up of severely malnourished children who participated in an intervention program. <i>Child Dev.</i> 1994 Apr;65(2 Spec No):428-39.
Pegboard	N/A	Brazil	mean age 9.4 years	Portuguese	Examination of Brazilian school children's performance on the Gardner Steadiness Test and the Purdue Pegboard.	Child sex, hand preference and age were linked to performance on the Pegboard, but social status and ethnic membership were not related to scores.	Brito et al	2002	Brito GN, Santos-Morales T, Developmental norms for the Gardner Steadiness Test and the Purdue Pegboard: A study with children of a metropolitan school in Brazil. <i>Brazilian Journal of Medical and Biological Research.</i> 2002, 35(8), 931-949.
Pegboard	N/A	Yemen Republic	school children		Examination of influence of asymptomatic parasites on school achievement in	Treatment of parasites was not associated with gains in cognitive functioning in a two-week period. Children without parasites performed better on fine	Al Serouri et al	2000	Al Serouri AW, Grantham-McGregor SM, Greenwood B, Costello A. Impact of

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					children in Yemen. Children with parasites were compared to children without on fine motor tests (including the pegboard) and other tests of cognitive functioning.	motor functioning than children with parasites.			asymptomatic malaria parasitaemia on cognitive function and school achievement of schoolchildren in the Yemen Republic. Parasitology, 2000, Oct;121 (Pt 4):337-45.
PPVT	N/A	China		Chinese	Speech therapy of 62 children with lingua-apical articulation disorder.	After 5 to 18 times of therapy, the effective rate of speech therapy reached 82.3%. (PPVT was used along with Gesell development scales, Wechsler Intelligence Scale for Preschool Children (WPPSI), and a speech test (unspecified in abstract)	Zhang et al.	2006	Zhang FH, Jin XM, Zhang YW, Wu H, Jiang F, Shen XM. Clinical . Clinical characteristics and speech therapy of lingua-apical articulation disorder. 2006 Mar;44(3):210-3.
Preschool Language Scale (PLS)	N/A	Seychelles Islands	66 months	Creole (French-based)	Evaluation of whether nonlinear effects of methylmercury exposure were present, using scores on six tests administered to cohort children in the Seychelles Child Development Study.	Overall there was no clear evidence for consistent (across the entire range of exposure levels) adverse effects of exposure on the six developmental outcomes. Further nonlinear modeling of these data may be appropriate, but there is also the risk of fitting complex models without a clear biological rationale.	Axtell et al.	2000	Axtell CD, Cox C, Myers GJ, et al. Association between methylmercury exposure from fish consumption and child development at five and a half years of age in the Seychelles Child Development Study: an evaluation of nonlinear relationships. Environ Res. Oct 2000;84(2):71-80.
Preschool Language Scale (PLS)	N/A	Seychelles Islands	66 months	Creole (French-based)	Study of the association between methylmercury (MeHg) exposure and the developmental outcomes of children in the Republic of Seychelles at 66 months of age.	No adverse outcomes at 66 months were associated with either prenatal or postnatal MeHg exposure. CONCLUSION: In the population studied, consumption of a diet high in ocean fish appears to pose no threat to developmental outcomes through 66 months of age.	Davidson et al.	1998	Davidson PW, Myers GJ, Cox C, et al. Effects of prenatal and postnatal methylmercury exposure from fish consumption on neurodevelopment: outcomes at 66 months of age in the Seychelles Child Development Study. Jama. Aug 26 1998;280(8):701-707.
Preschool Language Scale (PLS)	N/A	Seychelles Islands	66 months	Creole (French-based)	Study to examine association between prenatal or postnatal methylmercury exposure and child development.	A consistent pattern of effect modification (EM) of social and environmental factors has not been observed, suggesting that the results may be due to chance.	Davidson et al.	2004	Davidson PW, Myers GJ, Shamlaye C, Cox C, Wilding GE. Prenatal exposure to methylmercury and child development: influence of social factors. Neurotoxicol Teratol. Jul-Aug 2004;26(4):553-559.
Preschool Language Scale (PLS)	N/A	Seychelles Islands	66 months	Creole (French-based)	Study testing the hypothesis that prenatal exposure to low concentrations of methylmercury from a maternal diet high in fish is	The results indicated that mercury exposure was negatively associated with four endpoints (the McCarthy General Cognitive Index and Perceptual Performance subscale and The Preschool Language Scale Total Language and Auditory Comprehension	Myers et al.	1997	Myers GJ, Davidson PW, Shamlaye CF, et al. Effects of prenatal methylmercury exposure from a high fish diet on developmental milestones in

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					related to the child's developmental outcome.	subscale). After normalizing the data by removal of a small number of outliers or highly influential scores, the mercury effects were no longer significant except for auditory comprehension.			the Seychelles Child Development Study. Neurotoxicology. 1997;18(3):819-829.
Shoklo Neurological and Developmental Tests	N/A	Thailand	9 to 36 months (neurological) and 3 to 12 months (developmental)	English and Thai	To develop a test suitable for assessing acute neurological disturbances caused by such diverse effects as infections, drugs, or toxins, to follow the evolution of the signs tested initially, and to evaluate later neurodevelopmental sequelae which might be caused by the same events.	Both tests showed good correlations with the Griffiths Developmental Scales, a standardized neurodevelopmental assessment.	Haataja et al.	2002	Haataja L, McGready R, Arunjerdja R, Simpson J, Mercuri E, Nosten F, and Dubowitz L. (2002). A new approach for neurological evaluation of infants in resource-poor settings. <i>Annals of Tropical Paediatrics</i> , 22, 355-368.
Stanford Binet	N/A	India	6 years	Not specified	Study to examine the mortality and morbidity in high risk infants after discharge from the hospital. Children assessed at 30 months on the Bayley Scales were assessed at age 6 on the intelligence quotient of the Stanford-Binet Scale.	In the 286 children who were assessed at 6 years, the incidence of borderline intelligence was 14.6% as compared to 5.6% in the controls ($p < 0.05$). Children who appear to have normal development in the third year, may show a high incidence (14.6%) of borderline intelligence at six years.	Chaudhari et al.	1997	Chaudhari S, Bhalerao MR, Vaidya U, Pandit A, Nene U. Growth and development of twins compared with singletons at ages one and four years. <i>Indian Pediatr</i> . Dec 1997;34(12):1081-1086.
Stanford Binet	N/A	India	4 years	Not specified	Study to compare the growth and development of twins with normal control singletons and also with matched 'high risk' singletons, at one and four years of age. Bayley Scales were used to assess the children at 1 year. At 4 years, the intelligence quotient was determined by the Stanford Binet Intelligence Scale.	At one year, the development of twins was within normal limits although the motor quotients were significantly lower than that of controls. At 4 years, the intelligence quotients of twins were well within normal limits.	Chaudhari et al.	2000	Chaudhari S, Kulkarni S, Pandit A, Deshmukh S. Mortality and morbidity in high risk infants during a six year follow-up. <i>Indian Pediatr</i> . Dec 2000;37(12):1314-1320.
Stanford Binet	N/A	Japan	Mean ages of 6 years and 8.2 years	Japanese	Study to assess autistic symptom differences between high-functioning atypical autism (HAA) and childhood autism (HCA).	HAA children were significantly less abnormal than the HCA children. Affect tended to be significantly milder in HAA than HCA. Anxiety reaction was significantly more abnormal in HAA than HCA. These findings may be useful to distinguish between HAA and HCA.	Kanai et al.	2004	Kanai C, Koyama T, Kato S, Miyamoto Y, Osada H, Kurita H. Comparison of high-functioning atypical autism and childhood autism by Childhood Autism Rating Scale-Tokyo version. <i>Psychiatry Clin Neurosci</i> . Apr 2004;58(2):217-221.
Stanford Binet	N/A	Not specified - China or	Not specified	Not specified	The Stanford Binet was used to test intelligence in children	A lot of mental retarded children were found in the iodine-deficiency areas, with most of them born	Fu et al.	1994	Fu LX, Chen ZH, Deng LQ. [Effects of iodine nutritional

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
		Taiwan			without iodine supplement during their fetal and infant periods, and in those born three to four years after the implementation of stable supply of iodized salt in areas with endemic cretinism and goiter.	before the implementation of iodine supplementation. The tested children living in the areas with endemic cretinism were followed-up for two years, and no improvement in children's intelligence could be seen. Impairment to children's intelligence development caused by iodine deficiency during their fetal and infant periods was irreversible.			status of fetuses, infants and young children on their intelligence development in the areas with iodine-deficiency disorders]. Zhonghua Yu Fang Yi Xue Za Zhi. Nov 1994;28(6):330-332.
Stanford Binet	N/A	Taiwan	4 and 5 years	Chinese	Study to compare the cognitive development in Taiwanese children who had been exposed prenatally to high levels of heat-degraded polychlorinated biphenyls (PCBs) with control children who were exposed to background levels	Cognitive development was measured from 1985 through 1990 using the Chinese versions of the Stanford-Binet test and the Wechsler Intelligence Scale for Children, Revised. The exposed children scored approximately 5 points lower on the Stanford-Binet test at the ages of 4 and 5 years and approximately 5 points lower on the Wechsler Intelligence Scale for Children, Revised, at the ages of 6 and 7 years. Children prenatally exposed to heat-degraded PCBs had poorer cognitive development than their matched controls. The effect persisted in the children up to the age of 7 years, and children born long after the exposure were still affected.	Chen et al.	1992	Chen YC, Guo YL, Hsu CC, Rogan WJ. Cognitive development of Yu-Cheng ("oil disease") children prenatally exposed to heat-degraded PCBs. <i>Jama</i> . Dec 9 1992;268(22):3213-3218.
Stanford Binet	N/A	Taiwan	2 to 12 (but abstract doesn't specify which measures were used at which ages)	Chinese	Study to determine the effect of prenatal exposure to PCBs on cognitive development in Yucheng children.	Chinese version of the Stanford-Binet IQ Test and other measures were used to assess the cognitive development of these children. The Yucheng children scored lower than control children on each of these methods of measurement between the ages of 2 and 12 years. Prenatal exposure to PCBs and their derivatives has long-term adverse effects on cognitive development in humans.	Lai et al.	2001	Lai TJ, Guo YL, Guo NW, Hsu CC. Effect of prenatal exposure to polychlorinated biphenyls on cognitive development in children: a longitudinal study in Taiwan. <i>Br J Psychiatry Suppl</i> . Apr 2001;40:s49-52.
Stanford Binet	N/A	Thailand	2 and 5 years	Thai	Study to detect newborns with congenital hypothyroidism (CH) by TSH screening and to treat the affected infants as early as possible.	The Stanford Binet or Vineland Social Maturity scale was used to evaluate the patient's intelligence between the ages of 2 and 5 years. RESULTS: From 1995 to 2000, 62,681 neonates were screened for CH. The average intelligence score was 93.8. The factor that significantly affected the intellectual outcome of the patients was associated with adverse conditions such as prematurity with complications, epilepsy and family disruption ($p < 0.05$).	Churesigaew et al.	2002	Churesigaew S, Ratrisawasdi V, Thaeramanophab S. Thyrotropin screening for congenital hypothyroidism in Queen Sirikit National Institute of Child Health, Thailand (during year 1995-2000). <i>J Med Assoc Thai</i> . Jul 2002;85(7):782-788.
Stanford Binet	IV	Thailand	children 2-5 years old (older than 5 included in the study as well)	Thai translation and adaption for use in Thailand by qualified psychologists.	"Language development in 32 preschool siblings (aged 2-6 years) of children with diagnosed autistic spectrum disorder (ASD) was compared with that of a control group of 28 typical preschool children. "	"Eight of the siblings had delayed language development, of whom three received a diagnosis of developmental language disorder (DLD) and one of ASD. The sibling with ASD and two of those with DLD were excluded; the remaining 29 siblings and the controls were administered the Stanford-Binet IV. Verbal IQs of siblings were not significantly different from the control group. Siblings of children with ASD associated with intellectual	Chuthapisith J, Ruangdaraganon N, Sombuntham T, Roongpraiwan R.	2007	Chuthapisith J, Ruangdaraganon N, Sombuntham T, Roongpraiwan R. Language development among the siblings of children with autistic spectrum disorder. <i>Autism</i> . 2007 Mar;11(2):149-60.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						impairment ('mental retardation' (MR) in Thailand) had significantly lower verbal IQ scores than siblings of children with ASD but without MR."			<u>Other child development tests used:</u> N/A
Stanford Binet	N/A	Turkey	3 and 5 years at beginning of study	Turkish	This study examines the effects of two different types of early enrichment: child-focused (center-based) and mother-focused (home-based).	"Although both interventions produced superior cognitive skills and school adjustment at the end of the program, follow-up assessments in Study 2 revealed that parent-focused intervention had numerous sustained effects in terms of school attainment, higher primary school grades and vocabulary scores, more favorable attitudes towards school, and better family and social adjustment, while most effects of center-based intervention had dissipated (with the notable exception of negative effects of custodial, as opposed to educational day care). It is concluded that home-based early enrichment through the mediation of the mother is a highly effective strategy with multiple positive outcomes in contexts of socioeconomic disadvantage."	Kagitcibasi C, Sunar D, Bekman S	2001	Kagitcibasi C, Sunar D, Bekman S. Long-term effects of early intervention: Turkish low-income mothers and children. <i>Applied Developmental Psychology</i> , 2001, 22: 333-361. <u>Other child development tests used:</u> triad of subtests from the Wechsler series (including subtests from WPPSI and WISC-R); Children's Embedded Figures Test (CEFT) adapted for Turkish use; tests of academic achievement in Turkish, and grades from report cards
Strengths and Difficulties Questionnaire	N/A	Bangladesh	4 to 16 years	Not specified	Study to examine the potential suitability of the SDQ for detecting child psychiatric problems in Bangladesh.	SDQ scores distinguished well between community and clinic samples, and also between children with different psychiatric diagnoses in the clinic sample. Predictions based on multi-informant SDQs potentially provide a cheap and easy method for detecting children in the developing world with significant mental health problems. The potential effectiveness of any such screening programme should be evaluated on a broad range of children, using both international and culture-specific assessments.	Mullick et al.	2001	Mullick MS, Goodman R. Questionnaire screening for mental health problems in Bangladeshi children: a preliminary study. <i>Soc Psychiatry Psychiatr Epidemiol.</i> Feb 2001;36(2):94-99.
Strengths and Difficulties Questionnaire	N/A	Brazil, Pakistan, Canada, Bangladesh, Israel, Yemen, Thailand, Australia	4 to 16 years	Portuguese, Urdu, English, not specified (for Bangladesh, Israel, Yemen, Thailand)	Review article discussing application of SDQ in non-European countries. It summarizes some of these SDQ experiences by presenting a selection of projects that have either psychometrically evaluated this questionnaire, applied it to screen for behavior disorders, or employed its versions as research tools.	Across a huge variety of cultures and languages, experience gained with the SDQ in other continents has supported European evidence of good psychometric properties and clinical utility of this questionnaire.	Woerner et al.	2004	Woerner W, Fleitlich-Bilyk B, Martinussen R, et al. The Strengths and Difficulties Questionnaire overseas: evaluations and applications of the SDQ beyond Europe. <i>Eur Child Adolesc Psychiatry.</i> 2004;13 Suppl 2:II47-54.
Strengths and Difficulties Questionnaire	N/A	Democratic Republic of Congo	7 to 9 years	French	The SDQ was pilot-tested; the stability of the factor structure was examined and	The internal consistency was satisfactory on all of the SDQ scales. Using the 90th percentile, the cut-off scores were somewhat higher than the published	Kashala et al.	2005	Kashala E, Elgen I, Sommerfelt K, Tylleskar T. Teacher ratings of mental health among school

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					the reliability was evaluated.	cut-off scores in this younger sample. Poor nutrition, low socioeconomic status and illness were found to increase the risk for mental health problems and low school performance.			children in Kinshasa, Democratic Republic of Congo. Eur Child Adolesc Psychiatry. Jul 2005;14(4):208-215. (electronic version included)
Strengths and Difficulties Questionnaire	N/A	Israel	3 to 6 years	Not specified	Study to investigate the relationship between exposure to war trauma and behavioral and emotional problems among pre-school children.	The total number of traumatic events independently predicted total BCL and SDQ scores. Exposure to day raids and shelling of the children's houses by tanks were significantly associated with total behavioural and emotional problems scores. Direct and non-direct exposure to war trauma increases the risk of behavioural and emotional problems among pre-school children, which may present as non-specific psychopathology.	Thabet et al.	2006	Thabet AA, Karim K, Vostanis P. Trauma exposure in pre-school children in a war zone. Br J Psychiatry. Feb 2006;188:154-158.
Strengths and Difficulties Questionnaire	N/A	Israel	3, 6, 11, 16 years	Not specified	Study to establish the mental health profile among 322 Arab children living in the Gaza strip.	Western categories of mental health problems did not clearly emerge from the factor analysis. The main difference from western epidemiological studies appeared to operate in parents' perceptions of emotional problems in pre-school children.	Thabet et al.	2000	Thabet AA, Stretch D, Vostanis P. Child mental health problems in Arab children: application of the strengths and difficulties questionnaire. Int J Soc Psychiatry. Winter 2000;46(4):266-280.
Strengths and Difficulties Questionnaire	N/A	Pakistan	4 to 16 years	Urdu	Study to test the validity of the Urdu version of the SDQ.	The Urdu version was able to discriminate between the study groups. The mean difficulties score was higher in the case group (mean = 21.7) than the comparison group (mean = 14.4), the difference (mean difference = 7.3) being statistically significant ($p < 0.01$, 95% CI = 5.3, 9.3) with an effect size of 1.1. Adequate sensitivity ($> \text{or} = 61\%$) was obtained for each type of psychiatric disorder in the case group using the abnormal category of the SDQ symptom scores as a cutoff.	Samad et al.	2005	Samad L, Hollis C, Prince M, Goodman R. Child and adolescent psychopathology in a developing country: testing the validity of the strengths and difficulties questionnaire (Urdu version). Int J Methods Psychiatr Res. 2005;14(3):158-166
Stroop Test	N/A	Indonesia	8-9 years and 11-13 years	Javanese	Study investigated the association between helminth infection and cognitive ability in school children.	No associations were found between presence of hookworm and performance on Stroop test.	Sakhti et al.	1999	Sakti H, Nokes C, Hertanto WS, Hendratno S, Hall A, Bundy DA, Satoto. Evidence for an association between hookworm infection and cognitive function in Indonesian school children. Trop Med Int Health. 1999 May;4(5):322-34.
Stroop Test	N/A	Jamaica	9-12 years	Not specified	A double-blind placebo trial was conducted to determine the effect of moderate to high loads of <i>Trichuris trichiura</i> (whipworm) infection on the cognitive functions of 159 school children (age 9-12 years) in	The expulsion of worms led to a significant improvement in tests of auditory short-term memory (P less than 0.02; P less than 0.01), and a highly significant improvement in the scanning and retrieval of long-term memory (P less than 0.001). After 9 weeks, treated children were no longer significantly different from an uninfected Control group in these three tests of cognitive function. The	Nokes et al	1992	Nokes C, Grantham-McGregor SM, Sawyer AW, Cooper ES, Robinson BA, Bundy DA. Moderate to heavy infections of <i>Trichuris trichiura</i> affect cognitive function in Jamaican school children. Parasitology. 1992 Jun;104 (Pt 3):539-47.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					Jamaica	removal of <i>T. trichiura</i> was more important than <i>Ascaris lumbricoides</i> in determining this improvement. The results suggest that whipworm infection has an adverse effect on certain cognitive functions which is reversible by therapy.			
Stroop Test	Stroop/Pegboard	Tanzania	9-14 years	Kiswahili	Investigated the association between hookworm and cognitive function in Tanzania.	Children with hookworm scored lower on two tests of cognitive function, but were no different than children without hookworm on either the Stroop Test or the Pegboard.	Partnership for Child Development (Jukes et al)	2002	Jukes MC, Nokes CA, Alcock KJ, Lambo JK, Kihamia C, Ngorosho N, Mbise A, Lorri W, Yona E, Mwanri L, Baddeley AD, Hall A, Bundy DA; Partnership for Child Development. Heavy schistosomiasis associated with poor short-term memory and slower reaction times in Tanzanian schoolchildren. <i>Trop Med Int Health</i> . 2002 Feb;7(2):104-17.
Stroop Test	N/A	Turkey	children 4-5 years old (older than 5 included in the study as well)	Not specified in abstract	"The aim of this study was to examine the neurocognitive functions of the frontal lobe in parents of autistic children."	"Mothers of children with autism performed better than the control group mothers on the executive function measures of WCST. There were no group differences in Stroop Test measures of attention and inhibition, or in the verbal and performance intelligence subtests of WAIS. Fathers of children with severe autistic symptoms performed better on some WAIS subtests compared to other; however, there were no significant differences in IQ between the parents in both groups. The results suggest that parents of autistic children could display different cognitive styles, but we did not observe any distinctive cognitive profile pertaining to frontal lobe functions. The cognitive ability of parents of autistic children and its neurobiological basis should be further investigated."	Baykara B, Gencer O, Ilkin Z, Miral S.	2008	Baykara B, Gencer O, Ilkin Z, Miral S. [Neurocognitive features of the frontal lobe in parents of autistic children] <i>Turk Psikiyatri Derg</i> . 2008 Fall;19(3):225-34. <u>Other child development tests used:</u> Wisconsin Card Sorting Test (WCST), Wechsler Adult Intelligence Test (WAIS)
Stroop Test	N/A	United States	6 to 12 years	Spanish	Tested the Stroop Test in Spanish.	Determined developmental norms for testing in Spanish.	Armengol	2002	Armengol, GC. Stroop test in Spanish: children's norms. <i>Clinical Neuropsychology</i> , 2002 Feb;16(1):67-80.
Test de Desarrollo Psicomotor (TEPSI)	N/A	Chile	3 to 4 years	Spanish	Study to describe the psychomotor development, environmental stimulation, and socioeconomic condition of preschool children attending three educational institutions in Temuco, Chile.	Highest statistical correlation between psychomotor development and environmental stimulation when comparing all three parameters across sample.	Doussoulin Sanhueza A	2006	Doussoulin Sanhueza A. (2006). Psychomotor development, environmental stimulation, and socioeconomic level of preschoolers in Temuco, Chile. <i>Pediatric Physical Therapy : the Official Publication of the Section on Pediatrics of the American Physical Therapy Association</i> .

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Test de Desarrollo Psicomotor (TEPSI)	N/A	Chile	4 years	Spanish	To identify risk factors (RF) for developmental delay in pre-school children who belong to a Family Health Center serving a low-income urban/rural community, in the Metropolitan Region (North Area) of Santiago.	In 2003, 541 of 610 children under control at the Family Health Center had the TEPSI test due at age 4. SG included 76 children with developmental delay (abnormal test); CG included 83 children with normal development. 23% of the CG had some abnormal sub-test (usually involving language). RF for developmental deficit: male sex (p0.0006), poorer families (p0.0018), Family Health Center registration after 12 month of age (p 0.0075), not attending pre-school education (p 0.0075), suspected developmental deficit in previous evaluations (p 0.0013) and illiterate mother (p 0.018).	Schonhaut et al.	2005	18(2), 141-7. Schonhaut L, Rojas P, & Kaempffer R. (2005). Risk factors associated with developmental delay in pre-school children from a low income urban/rural community of Santiago, Chile. <i>Rev Chile Pediatrca</i> , 76(6), 589-598.
Wechsler Intelligence Scales for Children (WISC)	N/A	Australia	4 years	Not specified	Study to assess the relationship of birthweight to cognitive development.	The mean birthweight was 3386 g (SD: 517). There was a statistically significant association between birthweight and cognitive performance at age 2 years (adjusted deficit: 0.97 points per 100 g lighter; 95% CI: 0.4-1.5), but the magnitude of this association gradually decreased and became statistically non-significant at later childhood. The relationship between birthweight and cognitive development becomes progressively attenuated at increasing age. At older ages, socioenvironmental factors appear to play an increasingly important part in children's cognitive development.	Tong et al.	2006	Tong S, Baghurst P, McMichael A. Birthweight and cognitive development during childhood. <i>J Paediatr Child Health</i> . Mar 2006;42(3):98-103.
Wechsler Intelligence Scales for Children (WISC)	N/A	Bangladesh	10 years	Bengali	Cross-sectional investigation of intellectual function in 142 10-year-old children in Araihasar, Bangladesh, exposed to tube-well water containing manganese (Mn) and arsenic (As).	A dose-response relationship was found between exposure to Arsenic (As) in water and reduced intellectual function. Children with water As levels > 50 microg/L achieved significantly lower Performance and Full-Scale scores than did children with water As levels < 5.5 microg/L. The association was generally stronger for well-water As than for urinary As.	Wasserman et al.	2004	Wasserman GA, Liu X, Parvez F, Ahsan H, Factor-Litvak P, van Geen A, Slavkovich V, LoIacono NJ, Cheng Z, Hussain I, Momotaj H, Graziano JH. Water arsenic exposure and children's intellectual function in Araihasar, Bangladesh. <i>Environ Health Perspect</i> . 2004 Dec;112(17):A980.
Wechsler Intelligence Scales for Children (WISC)	N/A	Bangladesh	10 years	Bengali	Cross-sectional investigation of intellectual function in 142 10-year-old children in Araihasar, Bangladesh, exposed to tube-well water containing manganese (Mn) and arsenic (As).	After adjustment for sociodemographic covariates, water Mn was associated with reduced Full-Scale, Performance, and Verbal raw scores, in a dose-response fashion; the low level of As in water had no effect. In Bangladesh, some children are at risk for Mn-induced neurotoxicity.	Wasserman et al.	2006	Wasserman GA, Liu X, Parvez F, Ahsan H, Levy D, Factor-Litvak P, Kline J, van Geen A, Slavkovich V, LoIacono NJ, Cheng Z, Zheng Y, Graziano JH. Water manganese exposure and children's intellectual function in Araihasar, Bangladesh. <i>Environ Health Perspect</i> . 2006 Jan;114(1):124-9.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Wechsler Intelligence Scales for Children (WISC)	N/A	Belgium	7 - 11 years	Dutch	Study to assess the neuropsychological, academic and social-emotional profiles in Neurofibromatosis type 1 (NF1) children.	Nearly 50% (8/17) of the children showed learning disabilities. Isolated impaired literacy skills, particularly spelling problems, were most frequent (4/8). Nearly all NF1 children showed visual perceptual and executive dysfunctions. No correlation was found between the performances on the WISC-R, specific neuropsychological results, academic performances and behavioral problems.	Descheemaeker et al.	2005	Descheemaeker MJ, Ghesquiere P, Symons H, Fryns JP, Legius E. Behavioural, academic and neuropsychological profile of normally gifted Neurofibromatosis type 1 children. <i>J Intellect Disabil Res.</i> 2005 Jan;49(Pt 1):33-46.
Wechsler Intelligence Scales for Children (WISC)	N/A	Brazil	8 to 13 years	Portuguese	The intelligence quotient of obese children and adolescents was compared to that of non-obese children.	In spite of the fact that all the average IQ results presented consistently favored the eutrophic in relation to the obese, it is not possible to confirm one group's superiority over the other, due to the wide range of intervenient factors involved in the intelligence process.	Campos et al.	1996	Campos AL, Sigulem DM, Moraes DE, Escrivao AM, Fisberg M. Intelligent quotient of obese children and adolescents by the Wechsler scale. <i>Rev Saude Publica.</i> 1996 Feb;30(1):85-90.
Wechsler Intelligence Scales for Children (WISC)	N/A	Brazil	6 to 9 years	Portuguese	Study of association between early childhood diarrhea and parasitic infections and cognitive function.	Early diarrhea burdens correlated in pilot studies with impaired cognitive function using a McCarthy Draw-A-Design (P = 0.01; P = 0.017 when controlling for early helminth infections), Wechsler Intelligence Scale for Children coding tasks (P = 0.031), and backward digit span tests (P = 0.045).	Guerrant et al.	1999	Guerrant DI, Moore SR, Lima AA, Patrick PD, Schorling JB, Guerrant RL. Association of early childhood diarrhea and cryptosporidiosis with impaired physical fitness and cognitive function four-seven years later in a poor urban community in northeast Brazil. <i>Am J Trop Med Hyg.</i> 1999 Nov;61(5):707-13.
Wechsler Intelligence Scales for Children (WISC)	N/A	Brazil	4 to 7 years	Portuguese	Study to assess the potential long-term impact of early childhood diarrhea (in the first 2 years of life) on cognitive function in later childhood.	Wechsler Intelligence Scale for Children (WISC-III) Coding Tasks and WISC-III Digit Span (reverse and total) scores were significantly lower in the 17 children with a history of early childhood persistent diarrhea (PD; P < .05), even when controlling for helminths and maternal education.	Niehaus et al.	2002	Niehaus MD, Moore SR, Patrick PD, Derr LL, Lornitz B, Lima AA, Guerrant RL. Early childhood diarrhea is associated with diminished cognitive function 4 to 7 years later in children in a northeast Brazilian shantytown. <i>Am J Trop Med Hyg.</i> 2002 May;66(5):590-3.
Wechsler Intelligence Scales for Children (WISC)	N/A	Chile	Not specified	Spanish	Study used WISC to as genetic markers and total intelligence quotient (IQ) in children of both genders from Santiago, Chile.	Heterozygous boys for phosphoglucomutase 1 (PGM) and heterozygous girls for haptoglobin (Hp) had lower IQ than homozygotes. For ABO system, B girls had lower and B boys had higher IQ than children with other ABO phenotypes.	Valenzuela et al.	1998	Valenzuela CY, Pastene CS, Perez CM. Intelligence and genetic markers in Chilean children. <i>Biol Res.</i> 1998;31(2):81-92.
Wechsler Intelligence Scales for Children (WISC)	N/A	China	Not specified	Chinese	To study the intelligence of 22 patients with congenital velopharyngeal incompetence (CVPI).	More than half of the patients with CVPI showed Mental Retardation.	Fan et al.	2002	Fan ZJ, Wu YL, He BY, Wang GM. Intelligence evaluation of 22 patients with congenital velopharyngeal incompetence. <i>2002 Dec;11(4):314-5.</i> Shanghai Kou Qiang Yi Xue.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Wechsler Intelligence Scales for Children (WISC)	N/A	China	12 years	Chinese	Study to investigate the features of cognitive function in children with attention deficit/hyperactivity disorder (ADHD).	The levels of intelligence, memory, and attention in ADHD children are lower than those in normal children. ADHD children have deficiency in executive functions, such as selective inhibition, working memory and plan implementation.	Liu & Wang	2002	Liu Y, Wang Y. 2002 Mar 25;82(6):389-92. Cognitive functions of children with attention deficit/hyperactivity disorder [Article in Chinese. Zhonghua Yi Xue Za Zhi.
Wechsler Intelligence Scales for Children (WISC)	N/A	Colombia	5-16 years	Spanish	Study to obtain norms for the ENI in a Colombian population between 5 and 16 years of age. The WISC-R was used to obtain external validity.	The ENI could satisfy the existing need in the Spanish speaking world for neuropsychological tools with which to assess children and adolescents.	Rosseli Cock et al.	2004	Rosseli Cock M, Matute Villasenor E, Ardila Ardila A, Botero Gomez VE, Tangarife Salazar GA, Echevarria Pulido SE, Arbelaez Giraldo C, Mejia Quintero M, Mendez Losado LC, Villa Hurtado PC, Ocampo Agudelo P. Neuropsychological Assessment of Children: a test battery for children between 5 and 16 years of age. A Colombian normative study. Rev Neurol. 2004 Apr 16-30;38(8):720-31.
Wechsler Intelligence Scales for Children (WISC)	N/A	Denmark	8-9 years	Danish	Study to estimate prevalence rates of general and specific child psychiatric disorders in Danish children.	The overall estimated prevalence rate of child psychopathology was 11.8 % [95% confidence interval (CI): 8.8, 14.8]. Attention Deficit/Hyperactivity Disorder (ADHD) was found to be the most common specific child psychiatric disorder. The estimated prevalence rates were broadly comparable to prevalence rates found in other epidemiological studies.	Petersen et al.	2006	Petersen DJ, Bilenberg N, Hoerder K, Gillberg C. The population prevalence of child psychiatric disorders in Danish 8- to 9-year-old children. Eur Child Adolesc Psychiatry. 2006 Mar;15(2):71-8. Epub 2006 Feb 21.
Wechsler Intelligence Scales for Children (WISC)	N/A	Ecuador	Two lowest grades of a public school (but older than 9 years)	Spanish	Study to examine possible effects of prenatal pesticide exposure on blood pressure, neurological function, and neurobehavioral tests in school-aged children in an area where stunting is common.	Stunting was associated with a lower score on the WISC test only. Prenatal pesticide exposure may cause lasting neurotoxic damage and add to the adverse effects of malnutrition in developing countries. The effects differ from those due to acute pesticide exposure.	Grandjean et al.	2006	Grandjean P, Harari R, Barr DB, Debes F. Pesticide exposure and stunting as independent predictors of neurobehavioral deficits in Ecuadorian school children. Pediatrics. 2006 Mar;117(3):e546-56.
Wechsler Intelligence Scales for Children (WISC)	N/A	France	Not specified	French	Study of the influence of chronological age on verbal intelligence of persons with mental retardation.	The relationship between chronological age and verbal intelligence is valid for persons with moderate mental retardation as well as for those with mild mental retardation.	Facon & Facon-Bollengier	1999	Facon B, Facon-Bollengier T. Psychol Rep. 1999 Dec;85(3 Pt 1):857-62. Chronological age and performance of persons with mental retardation on verbal subtests of the Wechsler Intelligence Scale for Children—Revised, French version.
Wechsler Intelligence	Memory scale	India	children 5 years old (older than	Not specified in abstract	"As learning and memory in any field is achieved by	"The results showed that the abacus learners at the end of one and two years had a better visual and	Bhaskaran M, Sengottaiyan A,	2006	Bhaskaran M, Sengottaiyan A, Madhu S, Ranganathan V.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Scales for Children (WISC)			5 included in the study as well)		coordinating and analyzing the different sensory inputs, whether an abacus trainee would also improve the short-term memory as a whole was evaluated in our study."	auditory memory when compared to non-abacus learners."	Madhu S, Ranganathan V.		Evaluation of memory in abacus learners. Indian J Physiol Pharmacol. 2006 Jul-Sep;50(3):225-33. <u>Other child development tests used:</u> Mini mental state examination, Mann - Buitar visual memory screen for objects
Wechsler Intelligence Scales for Children (WISC)	Intelligence Scale for Children	India	children 5 years old (older than 5 included in the study as well)	Not specified in abstract; used subsections of the test that were familiar and considered culturally appropriate	To assess "children's intellectual function in relation to arsenic exposure."	"Stratifying urinary arsenic concentrations into tertiles, we found associations between arsenic and reductions in the adjusted scores of the vocabulary test (0, -0.14, -0.28; P for trend = 0.02), the object assembly test (0, -0.16, -0.24; P for trend = 0.03), and the picture completion test (0, -0.15, -0.26; P for trend = 0.02). These findings correspond to relative declines of 12% (95% confidence interval =-0.4% to 24%) in the vocabulary test, 21% (-0.8% to 42%) in the object assembly test, and of 13% (0.3% to 24%) in the picture completion test in the upper urinary arsenic tertile. However, we did not find evidence of an association between test results and arsenic water concentrations during pregnancy or childhood. Current arsenic concentrations in urine, which reflect all sources of recent exposure, including water and food, were associated with small decrements in intellectual testing in school-aged children in West Bengal. We did not see associations between long-term water arsenic concentrations and intellectual function."	von Ehrenstein OS, Poddar S, Yuan Y, Mazumder DG, Eskenazi B, Basu A, Hira-Smith M, Ghosh N, Lahiri S, Haque R, Ghosh A, Kalman D, Das S, Smith AH.	2007	von Ehrenstein OS, Poddar S, Yuan Y, Mazumder DG, Eskenazi B, Basu A, Hira-Smith M, Ghosh N, Lahiri S, Haque R, Ghosh A, Kalman D, Das S, Smith AH. Children's intellectual function in relation to arsenic exposure. Epidemiology. 2007 Jan;18(1):44-51. <u>Other child development tests used:</u> Raven Colored Progressive Matrices Test, Purdue Pegboard, Total Sentence Recall Test
Wechsler Intelligence Scales for Children (WISC)	N/A	India	9-14 years	Not specified	Study to assess the physical work capacity and cognition of underprivileged anemic schoolgirls in Vadodara in early adolescence as compared to their non-anemic counterparts.	Significantly lower scores in digit span and visual memory test were seen in anemic compared to non-anemic girls. The adverse impact of anemia remained after controlling for undernutrition (BMI). Anemia is likely to adversely affect physical work capacity and cognition in young adolescent girls undergoing pubertal development.	Sen et al.	2006	Sen A, Kanani SJ. Deleterious functional impact of anemia on young adolescent school girls. Indian Pediatr. 2006 Mar;43(3):219-26.
Wechsler Intelligence Scales for Children (WISC)	N/A	Iran	6 to 13 years	Not specified	A retrospective cohort study to evaluate the effects of maternal fasting during Ramadan on the intelligence quotient of their progeny at ages 4 - 13.	No significant differences were observed between the IQ scores of the two groups (children of mothers who fasted and children of mothers who didn't fast). Fasting during gestation did not adversely affect IQ of children whose mothers had fasted during Ramadan while being pregnant.	Azizi et al.	2004	Azizi F, Sadeghipour H, Shakholah B, Rezaei-Ghaleh N. Intellectual development of children born of mothers who fasted in Ramadan during pregnancy. Int J Vitam Nutr Res. 2004 Sep;74(5):374-80.
Wechsler Intelligence Scales for	N/A	Iran	6 years	Not specified	Study of the relationship between the Wechsler Intelligence Scale for	The scores on the two instruments correlated well.	Shahim et al.	1992	Shahim S. Correlations for Wechsler Intelligence Scale for Children--Revised and the

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Children (WISC)					Children--Revised (WISC--R) and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI).				Wechsler Preschool and Primary Scale of Intelligence for Iranian children. Psychol Rep. 1992 Feb;70(1):27-30.
Wechsler Intelligence Scales for Children (WISC)	N/A	Israel	14 to 16 years	Not specified	Study to assess the relationship between premature birth and cognitive development.	Results revealed that prematurely born adolescents scored lower than those born at term on all measures of cognitive performance.	Lubetzky et al.	1999	Lubetzky O, Weitzman A, Gilat I, Tyano S. Harefuah. Premature birth and cognitive functioning in adolescence. 1999 Nov 1;137(9):380-3, 430.
Wechsler Intelligence Scales for Children (WISC)	N/A	Italy	Not specified	Italian	Study to estimate potential early adverse effects on the Central Nervous System (CNS) due to very low exposure to inorganic lead.	The blood lead concentration was inversely and significantly associated with IQ, with an extrapolated decline of 1.29 points in total IQ for each microgram/dl increase of blood concentration.	Carta P et al.	2003	Carta P, Carta R, Girei E, Aru G, Flore C, Ibba A, Scaduto MT, Cadeddu C, Cocco S, Caracoi S, Sanna Randaccio F. Cognitive and performance capacity among adolescents living near a lead and zinc smelter. G Ital Med Lav Ergon. 2003 Jul-Sep;25 Suppl(3):43-5.
Wechsler Intelligence Scales for Children (WISC)	N/A	Mexico	6 to 9 years	Spanish	Cross-sectional study to examine the effects of chronic exposure to lead (Pb), arsenic (AS) and undernutrition on the neuropsychological development of children.	Higher levels of AsU were significantly related to poorer performance on WISC-RM factors examining long-term memory and linguistic abstraction, while lower scores in WISC-RM factors measuring attention were obtained at increasing values of PbB. Results suggest that exposure to As and chronic malnutrition could have an influence on verbal abilities and long-term memory, while Pb exposure could affect the attention process even at low levels.	Calderon et al.	2001	Calderon J, Navarro ME, Jimenez-Capdeville ME, Santos-Diaz MA, Golden A, Rodriguez-Leyva I, Borja-Aburto V, Diaz-Barriga F. Exposure to arsenic and lead and neuropsychological development in Mexican children. Environ Res. 2001 Feb;85(2):69-76.
Wechsler Intelligence Scales for Children (WISC)	N/A	Mexico	6 - 10 years	Spanish	Study to examine whether prenatal lead exposure has a more powerful and lasting impact on child development than postnatal exposure.	Lead exposure around 28 weeks gestation is a critical period for later child intellectual development, with lasting and possibly permanent effects. There was no evidence of a threshold; the strongest lead effects on IQ occurred within the first few micrograms of BPb.	Schnaas et al.	2006	Schnaas L, Rothenberg SJ, Flores MF, Martinez S, Hernandez C, Osorio E, Velasco SR, Perroni E. Reduced intellectual development in children with prenatal lead exposure. Environ Health Perspect. 2006 May;114(5):791-7. National Institute of Perinatology, Mexico City, Mexico.
Wechsler Intelligence Scales for Children (WISC)	N/A	Netherlands	6-12 years	Dutch	Study to examine VIQ-PIQ differences in children with autism and children with PDD-NOS.	Children with PDD-NOS seemed to have a similar VIQ-PIQ profile as children with autism, and on the subtest level children with PDD-NOS showed some similarities to children with Asperger syndrome or autism. It was not possible to distinguish PDD-NOS from autism or Asperger syndrome by using IQ scores.	deBruin et al.	2006	de Bruin EI, Verheij F, Ferdinand RF. WISC-R Subtest But No Overall VIQ-PIQ Difference in Dutch Children with PDD-NOS. J Abnorm Child Psychol. 2006 Feb 17; [Epub ahead of print]
Wechsler	N/A	Peru	9 years	Spanish	Study to assess the effect of	Children with severe stunting in the second year of	Berkman et al.	2002	Berkman DS, Lescano AG,

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Intelligence Scales for Children (WISC)					stunting, diarrhoeal disease, and parasitic infections during infancy on cognitive function in late childhood.	life scored 10 points lower on the WISC-R test (95% CI 2.4--17.5) than children without severe stunting. Children with more than one episode of G lamblia per year scored 4.1 points (0.2--8.0) lower than children with one episode or fewer per year. Malnutrition in early childhood, indexed by stunting, and potentially G lamblia, are associated with poor cognitive function at age 9 years.			Gilman RH, Lopez SL, Black MM. Effects of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late childhood: a follow-up study. <i>Lancet</i> . 2002 Feb 16;359(9306):564-71.
Wechsler Intelligence Scales for Children (WISC)	N/A	Russia	6 and 7 years	Russian	Study of the influence of genetic effects, common or shared family environment and unique environment on the development of psychometric intelligence during the transition from preschool to school.	At age 6, genetic influences were much greater than those of shared environment but the magnitude of genetic influences decreased and the magnitude of shared environment influences increased substantially by age 7.	Malykh et al. (17)	2003	Malykh SB, Zyrianova NM, Kuravsky LS. Longitudinal genetic analysis of childhood IQ in 6- and 7-year-old Russian twins. <i>Twin Res</i> . 2003 Aug;6(4):285-91.
Wechsler Intelligence Scales for Children (WISC)	N/A	South Africa	Third grade children	Not specified	Study compared the performance of 21 black and 35 white third-grade South African children on the K-ABC and the Wechsler Intelligence Scale--Revised (WISC-R) at two schools for children with learning problems.	While the WISC-R Verbal and Full Scale IQ of black children were significantly lower than that of whites, there was no significant difference between these groups on the K-ABC. Support is provided for the usefulness of the K-ABC as a relatively nondiscriminatory alternative to the WISC-R for South African children.	Skuy et al.	2000	Skuy M, Taylor M, O'Carroll S, Fridjhon P, Rosenthal L. Performance of black and white South African children on the Wechsler Intelligence Scale for Children--Revised and the Kaufman Assessment Battery. <i>Psychol Rep</i> . 2000 Jun;86(3 Pt 1):727-37.
Wechsler Intelligence Scales for Children (WISC)	N/A	Spain	Not specified	Spanish	To analyze the psychological sequelae in long-term survivors of childhood cancer and to establish the relationship between the changes produced in both cognitive and emotional factors and diagnostic and therapeutic variables as well as in sensorial sequelae.	Intelligent quotient was within normal limits. Lower scores were related to cranial radiotherapy, age at diagnosis and relapses. Emotionally, the survivors coped successfully with cancer, depressive symptoms being more prevalent than in the general population and anxiety almost negligible.	Benedito Monleon et al.	2000	Benedito Monleon MC, Lopez Andreu JA, Serra Estelles I I, Harto Castano M, Gisbert Aguilar J, Mulas Delgado F, Ferris I Tortajada J. <i>An Esp Pediatr</i> . 2000 Dec;53(6):553-60. Psychological sequelae in longterm cancer survivors.
Wechsler Intelligence Scales for Children (WISC)	N/A	Spain and Canada	Not specified	English and Spanish	Study to examine the role of intelligence (IQ) in the definition of reading disabilities (RD) in languages with different orthographic systems.	There were some differences between the groups of Canadian children with RD but not between the groups of Spanish children. There were differences in reading tasks as a function of Performance IQ in English but not in Spanish. The differences in the role of IQ as a function of orthographic systems may relate to the greater significance of visual-orthographic as opposed to phonological processing in English.	Jimenez et al.	2003	Jimenez JE, Siegel LS, Lopez MR. The relationship between IQ and reading disabilities in English-speaking Canadian and Spanish children. <i>J Learn Disabil</i> . 2003 Jan-Feb;36(1):15-23.
Wechsler Intelligence Scales for	N/A	Taiwan	Chinese	6 and older	Study to examine cognitive outcomes for children who exposed to biphenyl in	The exposed children scored 3 points (P =.05) lower than control children for IQ; 3 points (P =.002) higher on the Child Behavior Checklist (an effect	Lai et al.	2002	Lai TJ, Liu X, Guo YL, Guo NW, Yu ML, Hsu CC, Rogan WJ. A cohort study of

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Children (WISC)					cooking oil.	size similar to the sex difference); and 6 points (P<.001) higher on the Rutter scale (3 times the sex difference)Prenatal exposure to these compounds produces long-lasting cognitive and behavioral damage, but there is some evidence of recovery.			behavioral problems and intelligence in children with high prenatal polychlorinated biphenyl exposure. Arch Gen Psychiatry. 2002 Nov;59(11):1061-6.
Wechsler Intelligence Scales for Children (WISC)	N/A	Thailand	Not specified	Not specified	Study to determine the efficacy of a seasoning powder fortified with iron, iodine, vitamin A and zinc served with noodles or rice consumed for school lunch on biochemical status and cognitive function in 567 rural NE Thai children.	Seasoning powder fortified with four micronutrients reduced the incidence of zinc and iodine deficiency and increased haemoglobin concentration over 31 weeks, while at the same time improving short term memory and attention, and thus may contribute to improved overall cognitive functioning over time.	Manger et al.	2004	Manger MS, Winichagoon P, Pongcharoen T, Gorwachirapan S, Boonpradern A, McKenzie J, Gibson RS. Multiple micronutrients may lead to improved cognitive function in NE Thai schoolchildren. Asia Pac J Clin Nutr. 2004;13(Suppl):S46.
Wechsler Intelligence Scales for Children (WISC)	N/A	Turkey	Over 6 years old	Not specified	Study to examine psychometric outcomes of patients with meningomyelocele.	Fifty-four percent of the hydrocephalic patients had a normal development or IQ. Psychometric tests were normal in 76% of those without hydrocephalus.	Mirzai et al.	1998)	Mirzai H, Ersahin Y, Mutluer S, Kayahan A. Outcome of patients with meningomyelocele: the Ege University experience. Childs Nerv Syst. 1998 Mar;14(3):120-3.
Wechsler Intelligence Scales for Children (WISC)	N/A	Turkey	1st grade children	Turkish	Case-control study to investigate the relationship between a variety of physical, environmental and sociodemographic factors, and primary school performance.	The chance of having an above average full scale WISC-R score was lower among children whose either parent smoked cigarettes, whose height-for-age percentile was below 10, and whose mothers' education was less than 8 years.	Ozmert et al.	2005	Ozmert EN, Yurdakok K, Soysal S, Kulak-Kayikci ME, Belgin E, Ozmert E, Laleli Y, Saracbası O. Relationship between physical, environmental and sociodemographic factors and school performance in primary schoolchildren. J Trop Pediatr. 2005 Feb;51(1):25-32. Epub 2004 Dec 15.
Wechsler Intelligence Scales for Children (WISC)	N/A	Venezuela	Spanish+C162	Mean age 7.64 years (both WISC and WSSPE were used; instrument used for each age group not specified in abstract).	Study to estimate the rate of prevalence of attention deficit hyperactivity disorder (ADHD) in school aged children.	The estimated prevalence of ADHD was 10.15%, and a figure of 2.03% was obtained for the hyperactive type, 0.51% for the inattentive subtype, and 7.61% for the combined subtype. Prevalence was higher for males (ratio 3:1). Performance in the cognitive and academic measures was within the average interval. The prevalence of ADHD is similar to that reported in other countries.	Montiel-Nava et al.	2003	Montiel-Nava C, Pena JA, Montiel-Barbero I. Epidemiological data about attention deficit hyperactivity disorder in a sample of Marabino children. Rev Neurol. 2003 Nov 1-15;37(9):815-9.
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	5 European countries: Belgium, Denmark, Greece,	5 years	English, Dutch, Danish, Swedish, or Greek	Study to shed more light on the cognitive and motor development of 5-year-old intracytoplasmic sperm injection (ICSI-conceived)	The motor and cognitive development of ICSI-conceived children is very similar to that of naturally conceived children. However, demographic factors such as maternal educational level and maternal age at the time of the birth might	Ponjaert-Kristoffersen et al.	2005	Ponjaert-Kristoffersen I, Bonduelle M, Barnes J, et al. International collaborative study of intracytoplasmic sperm injection-conceived, in vitro

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
		Sweden, and the United Kingdom			children.	play different roles in the cognitive development of ICSI and in vitro fertilization (IVF)-conceived children, compared with NC children.			fertilization-conceived, and naturally conceived 5-year-old child outcomes: cognitive and motor assessments. <i>Pediatrics</i> . Mar 2005;115(3):e283-289.
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	Brazil	Pre-school age (exact age not specified)	Portuguese	Study to examine the cognitive development of school-aged children born preterm and with very low birthweight.	The children who entered this study had borderline intellectual functioning at the moment of the evaluation. Results indicate that they may face learning difficulties at school, thus requiring adequate stimuli that should be provided by the family and the school.	Meio et al.	2003	Meio MD, Lopes CS, Morsch DS. Prognostic factors for cognitive development of very low birth weight premature children. <i>Rev Saude Publica</i> . Jun 2003;37(3):311-318.
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	Brazil	4 and 5 years	Portuguese	Study to describe the cognitive development of a population of premature newborns and to assess possible prognostic factors for abnormalities.	Compared to the literature, these children showed a more severe cognitive development impairment. Small for gestational age, abnormal cerebral ultrasound exam and male sex were prognostic factors for worse outcome.	Meio et al.	2004	Meio MD, Lopes CS, Morsch DS, et al. [Pre-school cognitive development of very low birth weight preterm children]. <i>J Pediatr (Rio J)</i> . Nov-Dec 2004;80(6):495-502.
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	Brazil		Portuguese	Study to verify inter-observer reliability of the test for a study on preschool-age cognitive development in a cohort of very low birth weight (VLBW) premature children from the Fernandes Figueira Institute (IFF).	Application of the WPPSI-R scale in the study of cognitive development of VLBW premature children at IFF proved adequate.	Meio et al.	2001	Meio MD, Lopes CS, Sichiari R, Morsch DS. [Reliability of the WPPSI-R test in the evaluation of cognitive development in preschool children]. <i>Cad Saude Publica</i> . Jan-Feb 2001;17(1):99-105.
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	Brazil	children 4-5 years old	Not specified in abstract or methods	"To assess cognitive and behavioral development at preschool age of children born preterm and with low birth weight and raised in a developing country."	"WPPSI scores were: total intelligence quotient (IQ) 88.00+/-16.96, verbal IQ 89.72+/-16.72, and executive IQ 88.12+/-15.71 for the group with less than 1,500 grams; and total IQ 91.11+/-14.73, verbal IQ 93.36+/-12.65, and executive IQ 90.20+/-16.06 for the group between 1,500 and 2,500 grams. The best scores were obtained in tests that evaluated capacity of abstraction and symbolization, picture completion and common perception, in which only 5% and 6.3% of the children had abnormal results, respectively. The lowest scores were obtained in tests that evaluated visual-motor coordination and flexibility-speed of reasoning, in which 27.5% and 16.3% of the children had abnormal results, respectively. A total of 32.5% had abnormal results in the arithmetic tests. Behaviors related to attention deficit/hyperactivity disorder (ADHD) were observed in 48% of the sample. Low score in the Bayley Mental Scale and abnormal result in the Denver test were significantly correlated to ADHD (p = 0.017 and p = 0.004). Abnormal results in the Bayley Mental Scale (p < 0.001), Denver test (p < 0.001) and neurological examination (p = 0.002)	do Espírito Santo JL, Portuguez MW, Nunes ML	2009	do Espírito Santo JL, Portuguez MW, Nunes ML. Cognitive and behavioral status of low birth weight preterm children raised in a developing country at preschool age. <i>J Pediatr (Rio J)</i> . 2009 Jan-Feb;85(1):35-41. Epub 2009 Jan 9. <u>Other child development tests used:</u> Denver, Conners' Parent Rating Scale-Revised (CPRS-R, adapted and validated in Portuguese), Bayley

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						were associated with lower IQ. The results revealed an increased incidence of behavioral and cognitive disorders at preschool age."			
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	R	Brazil	WPPSI-R given at age 5; children younger than 42 months during the first 6 months of 1999 were eligible to participate in the cognitive study sub-sample, ages ranged from 4.2 to 42.1 months	Translated into Portugese and previously applied in Brazil	"This study aimed to examine the relative contributions of both proximal and distal risk factors on child cognitive development, by breaking down the possible causal pathways through which poverty affects cognition."	"Unfavourable socioeconomic conditions, poorly educated mother, absent father, poor sanitary conditions at home and in the neighbourhood and low birth weight were negatively associated with cognitive performance at five years of age, while strong positive associations were found with high levels of domestic stimulation and nursery school attendance. Children's cognitive development in urban contexts in developing countries could be substantially increased by interventions promoting early psychosocial stimulation and preschool experience, together with efforts to prevent low birth weight and promote adequate nutritional status."	Santos DN, Assis AM, Bastos AC, Santos LM, Santos CA, Strina A, Prado MS, Almeida-Filho NM, Rodrigues LC, Barreto ML.	2008	Santos DN, Assis AM, Bastos AC, Santos LM, Santos CA, Strina A, Prado MS, Almeida-Filho NM, Rodrigues LC, Barreto ML. Determinants of cognitive function in childhood: a cohort study in a middle income context. BMC Public Health. 2008 Jun 6;8:202. <u>Other child development tests used:</u> N/A
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	R	China	children, age not specified in abstract	Chinese version	"The aim of the study was to evaluate the long-term effects of neonatal short bowel syndrome on cognitive functions during development."	"Eight of 9 patients were followed up except for one patient who died in a car accident at the age of three. All patients had been weaned off parenteral nutrition for more than 2 years. The average residual small bowel length was 58.1 cm (range 35-70 cm), and the mean parenteral nutrition (PN) duration was 73.1 days (43-147 days). The mean duration of the period without PN was 7.4 years (range 2.1-17.1 years). Weight, height and BMI for age were normal in 7 children except for 1 child, who was overweight. Hemoglobin and albumin concentrations were normal in all 8 patients. Evaluation of cognitive development showed normal results for all 8 patients while a verbal/performance discrepancy was found in 2. Patients with neonatal SBS who were weaned off PN for more than 2 years were found to have normal growth and cognitive development during this long-term follow-up. There was no evidence for a strong correlation between SBS and nutritional/cognition disorder. Longer term and controlled studies with a larger sample size are warranted."	Huang J, Cai W, Tang Q, Feng Y, Tao Y, Wang Y, Wu J.	2008	Huang J, Cai W, Tang Q, Feng Y, Tao Y, Wang Y, Wu J. Long-term cognitive functions in neonatal short bowel syndrome patients. Eur J Pediatr Surg. 2008 Apr;18(2):89-92. <u>Other child development tests used:</u> chinese versions of WAIS-R and WISC-R, BSID assessment on patients less than 4 years old.
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	China	children 48-72 months (younger infants included in the study but given Gesell Developmental Schedule)	Not specified in abstract or methods	"To investigate the impact of asymptomatic congenital CMV infection on physical and intellectual development of children during the first 6 years of life in Qinba mountain area."	"Either in neonatal or in infant period, no significant difference was noted between the asymptomatic congenital CMV infection children and the controls in average weight, height and head circumference (both p>0.05). The intellectual development was disproportion in asymptomatic congenital infected children. Compared with the control group, both global development quotient (DQ) and full-scale intelligence quotient (IQ) scores of	Zhang XW, Li F, Yu XW, Shi XW, Shi J, Zhang JP.	2007	Zhang XW, Li F, Yu XW, Shi XW, Shi J, Zhang JP. Physical and intellectual development in children with asymptomatic congenital cytomegalovirus infection: a longitudinal cohort study in Qinba mountain area, China. J Clin Virol. 2007 Nov;40(3):180-5.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						asymptomatically infected children were worse (t=2.19, p=0.031; t=2.48, p=0.015), especially on language DQ scores (t=3.25, p=0.002) and verbal IQ scores (t=3.88, p=0.000). However, the incidence rates of mental retardation (DQ/IQ<70) were similar in two groups (chi(2)=1.03, p>0.05). Although asymptomatic congenital CMV infection did not have significant influence on the neonatal physical development or incidence of mental retardation later in life, it is obviously an important factor correlating with long-time cognitive outcomes, especially on the development of language. It is necessary to survey CMV congenital infection and monitor the early intellectual development of children with asymptomatic congenital CMV infection in this area."			<u>Other child development tests used:</u> Gesell Developmental Schedule used to assess the development quotient of infants 18-36 months (table also lists infants 10-36 months)
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	China	children 3-5 years old (older than 5 included in the study as well)	Not specified in abstract	"To explore the clinical characteristics and speech therapy of 62 children with lingua-apical articulation disorder. "	"The 62 cases with the apical articulation disorder were classified into four groups. The combined pattern of the articulation disorder was the most common (40 cases, 64.5%), the next was apico-dental disorder (15 cases, 24.2%). The third was palatal disorder (4 cases, 6.5%) and the last one was the linguo-alveolar disorder (3 cases, 4.8%). The substitution errors of velar were the most common (95.2%), the next was omission errors (30.6%) and the last was absence of aspiration (12.9%). Oral motor dysfunction was found in some children with problems such as disordered joint movement of tongue and head, unstable jaw, weak tongue strength and poor coordination of tongue movement. Some children had feeding problems such as preference of eating soft food, keeping food in mouths, eating slowly, and poor chewing. After 5 to 18 times of therapy, the effective rate of speech therapy reached 82.3%. The lingua-apical articulation disorders can be classified into four groups. The combined pattern of the articulation errors is the most common one. Most of the apical sounds are replaced by velar sounds. The speech localization method is very useful in the therapy of apical articulation disorder. For children with feeding problems and oral motor dysfunction, it is needed to improve food texture and administer oral motor skill training."	Zhang FH, Jin XM, Zhang YW, Wu H, Jiang F, Shen XM.	2006	Zhang FH, Jin XM, Zhang YW, Wu H, Jiang F, Shen XM. [Clinical characteristics and speech therapy of lingua-apical articulation disorder] Zhonghua Er Ke Za Zhi. 2006 Mar;44(3):210-3. <u>Other child development tests used:</u> Peabody Picture Vocabulary Test (PPVT), Gesell development scales (Gesell)
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	France	5 years	French	Study to assess the impact of prematurity, fetal hypotrophy and familial environment on the neurodevelopmental	The five-year neurologic outcome of the children born prematurely in this regional study is similar to the results observed in regional studies conducted in Europe: 13.4% of the survivors have cerebral palsy, and the cognitive functions of the children with no	Burguet et al.	2000	Burguet A, Monnet E, Roth P, et al. Neurodevelopmental outcome of premature infants born at less than 33 weeks of gestational age and not cerebral

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					performances of very premature infants without cerebral palsy at the age of five years.	cerebral palsy are significantly lower than the control group (born at term).			palsy at the age of 5 years. Arch Pediatr. Apr 2000;7(4):357-368. (electronic version included)
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	Iran	4 to just under 6 years	Not specified	A retrospective cohort study to evaluate the effects of maternal fasting during Ramadan on the intelligence quotient of their progeny at ages 4 - 13.	No significant differences were observed between the IQ scores of the two groups (children of mothers who fasted and children of mothers who didn't fast). Fasting during gestation did not adversely affect IQ of children whose mothers had fasted during Ramadan while being pregnant.	Azizi et al.	2004	Azizi F, Sadeghipour H, Shiahkoh B, Rezaei-Ghaleh N. Intellectual development of children born of mothers who fasted in Ramadan during pregnancy. Int J Vitam Nutr Res. 2004 Sep;74(5):374-80.
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	Iran	6 years	Not specified	Study of the relationship between the Wechsler Intelligence Scale for Children--Revised (WISC--R) and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI).	The scores on the two instruments correlated well.	Shahim et al.	1992	Shahim S. Correlations for Wechsler Intelligence Scale for Children--Revised and the Wechsler Preschool and Primary Scale of Intelligence for Iranian children. Psychol Rep. 1992 Feb;70(1):27-30.
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	R	Mexico	children 5 years old	Not specified in abstract	To compare "the results of the interwave intervals of brainstem auditory evoked potentials (BAEP) at 80 dB SPL (sound pressure level) and the latencies of the waves of long latency auditory evoked potentials (LLAEP) (vertex) in 5-year-old children with articulatory defects for /l/, /r/, /rr/ and /s/, and in controls."	"Data were analysed using SPSS 12, and descriptive statistics and Student's t test were carried out to appraise the differences between the two groups. No differences were found between the control group and the study group. It is important to determine whether children with articulatory defects have alterations in the auditory receptor or in their central auditory function. It can be concluded that the variables analysed in the two groups behave in the same way; no significant differences were found, which suggests that neither the auditory receptor nor the central auditory function are affected in the study group. Further studies are to be conducted to investigate the meaning of the difference in central nervous conduction between the two differences in the study group."	Uribe-Escamilla R, Penaloza-Lopez YR, Durand-Rivera JA, Garcia-Pedroza F, Villarruel C.	2006	Uribe-Escamilla R, Penaloza-Lopez YR, Durand-Rivera JA, Garcia-Pedroza F, Villarruel C. [Brainstem auditory evoked potentials and vertex potentials in 5-year-old children with defective articulation of the phonemes /l/, /r/, /rr/ and /s/]. Rev Neurol. 2006 Sep 16-30;43(6):330-4. <u>Other child development tests used:</u> N/A
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	Mexico	possibly some children 0-5 but age range not specified in abstract, 9 children aged 6.8 years (SD 2)	Not specified in abstract	"We undertook this study to demonstrate the visuomotor alterations, intelligence level and depression changes in children with recurrent strabismus. "	"We included nine children aged 6.8 years (SD 2). Overfunction of oblique muscles and dissociated strabismus were related to recurrence of strabismus. Stereovision was present in five cases previous to recurrence (rate: 170 sec of arc), and three lost this with recurrence of strabismus. Psychological test determined difficulties in socialization and signs of aggression, including data on depression and "dullness." Bender Test showed relevant defects in fine hand movement, level: 5.4 (SD 1.7). Santucci evaluation for Bender was 3.83 (SD 2.1). Correlation coefficient between values was significant for Santucci evaluation and stereovision (0.89). Global Intelligence Coefficient was 88.1 (SD	Moguel-Ancheita S, Ramírez-Sibaja S, Castellanos-Pérez Bolde C, Orozco-Gómez LP.	2008	Moguel-Ancheita S, Ramírez-Sibaja S, Castellanos-Pérez Bolde C, Orozco-Gómez LP. [Study of depression and sensorial functions in children with strabismus. First phase.] Cir Cir. 2008 Mar-Apr;76(2):101-7. <u>Other child development tests used:</u> Intelligence Level of Weschler, WISC-R, Human Figure Test, Lauretta Bender Visuomotor

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
						12), which was subnormal and poorer in executive function (84). We have demonstrated relevant alterations in visuomotor abilities in patients with strabismus, especially related to stereovision deficiency, effect on learning, intelligence and depression."			Test
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	Pakistan	15 days to 72 months	Not specified	Study to investigate the association of parental consanguinity and delayed development in terms of gross motor, fine motor, speech and social aspects in their children.	Univariate analysis showed high significant Odds ratios for all areas of developmental delay (viz. gross and fine motor, speech and social development), between cases and controls (p < .001). At the multivariate analysis level however, the results showed no increased risk of parental consanguinity on delayed gross and fine motor, speech and social development in their children.	Ibrahim et al.	2001	Ibrahim S, Habib Z, Hyder S, Azam IS, Ahmed R. Parental [correction of Perinatal] consanguinity: a risk factor for developmental delay in Pakistani children. J Pak Med Assoc. Dec 2001;51(12):418-422.
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	Venezuela	Mean age 7.64 years (both WISC and WSSPE were used; instrument used for each age group not specified in abstract).	Spanish	Study to estimate the rate of prevalence of attention deficit hyperactivity disorder (ADHD) in school aged children.	The estimated prevalence of ADHD was 10.15%, and a figure of 2.03% was obtained for the hyperactive type, 0.51% for the inattentive subtype, and 7.61% for the combined subtype. Prevalence was higher for males (ratio 3:1). Performance in the cognitive and academic measures was within the average interval. The prevalence of ADHD is similar to that reported in other countries.	Montiel-Nava et al.	2003	Montiel-Nava C, Pena JA, Montiel-Barbero I. Epidemiological data about attention deficit hyperactivity disorder in a sample of Marabino children. Rev Neurol. 2003 Nov 1-15;37(9):815-9.
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	R	Venezuela	children 4-5 years old (older than 5 included in the study as well)	Not specified in abstract	"The aim of this study was to collect information about the characterization of the disorder depending on the gender. "	"The results indicated the lack of significant differences between genders for the studied variables, ADHD boys and girls scored alike in the various behavioral and cognitive measures. The results presented describe homogeneity in the symptoms, demographic characteristics and neuropsychological functioning for children of both genders; suggesting a syndrome with the same criteria and independent of the gender."	Montiel-Nava C, Montiel-Barbero I, Peña JA.	2007	Montiel-Nava C, Montiel-Barbero I, Peña JA. [Clinical presentation of attention deficit/hyperactivity disorder as a function of the gender] Invest Clin. 2007 Dec;48(4):459-68. <u>Other child development tests used:</u> Conners' rating scales-Revised, Wechsler Intelligence Scale for Children-Third Edition (WISC-III) , Diagnostic Interview Schedule for Children-IV Version-Parents (DISC-IV).
Wechsler Preschool and Primary Scales of Intelligence (WPPSI)	N/A	Yugoslavia	3, 4, and 5 years	Serbo-Croatian and Albanian	Study to investigate associations between the timing of lead (Pb) exposure on early intelligence, at ages 3, 4, 5, and 7 years, of children whose mothers lived in a smelter town during pregnancy.	Elevations in both prenatal and postnatal BPb were associated with small decrements in young children's intelligence.	Wasserman et al.	2000	9. Wasserman GA, Liu X, Popovac D, et al. The Yugoslavia Prospective Lead Study: contributions of prenatal and postnatal lead exposure to early intelligence. Neurotoxicol Teratol. Nov-Dec 2000;22(6):811-818.

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
Wisconsin Card-Sorting Task	N/A	Colombia	5-12 years	Spanish	Investigation of the properties of the WCST within a Colombian population.	Results indicated that scores were not related to SES, age or gender, but were predictive of academic achievement.	Rosselli et al.	1993	Rosselli M, Ardila, A. Developmental norms for the Wisconsin Card-Sorting Task for children aged 5 to 12 years. Clinical Neuropsychologist, 1993, Vol 7., p. 145-154.
Wisconsin Card-Sorting Task	N/A	Taiwan	6-11 years	Mandarin Chinese and English	The main aims of this study were to (1) develop norms for the Wisconsin Card Sorting Test (WCST) in children in Taiwan; (2) explore the effect of sex, age, birth order, number of siblings, and parental education on WCST performance; and (3) make a comparison of WCST performance between children in Taiwan and the US. 219 6-11 yr olds participated in this study.	The results indicate that starting at the age of 6, children make rapid gains in the number of categories achieved and significantly reduce the number of perseverative errors with each subsequent year. It is concluded that the results of this comparison of developmental norms of school children in Taiwan and the US may facilitate the WCST as a clinical or research instrument in combination with other test procedures to assess aspects of cognitive and neuropsychological functioning of school children.	Shu et al.	2000	Shu B-C, Tien A, Lung FW, Chang YY. Norms for the Wisconsin Card Sorting Test in 6- to 11-year-old children in Taiwan. Clinical Neuropsychologist, 2000, 14(3), 275-286.
Woodcock-Johnson	N/A	Costa Rica	5 years	Spanish	Follow-up evaluation (at five years) of a group of Costa Rican children whose iron status and treatment were documented in infancy.	The mean (+/- SD) adjusted Woodcock-Johnson preschool cluster score for the children who had moderate anemia in infancy (n = 30) was 448.6 +/- 9.7, as compared with 452.9 +/- 9.2 for the rest of the children (n = 133) (P less than 0.01); the adjusted visual-motor integration score was 5.9 +/- 2.1, as compared with 6.7 +/- 2.3 (P less than 0.05). Children who have iron-deficiency anemia in infancy are at risk for long-lasting developmental disadvantage as compared with their peers with better iron status.	Lozoff et al.	1991	Lozoff B, Jimenez E, Wolf AW. Long-term developmental outcome of infants with iron deficiency. N Engl J Med. 1991 Sep 5;325(10):687-94.
Woodcock-Johnson	N/A	no applicable articles in pubmed search	N/A	N/A	N/A	N/A	N/A	N/A	N/A <u>Other child development tests used:</u> N/A
Woodcock-Johnson	N/A	Not specified	6-12 years	Spanish	Study to assess cognitive disorders in epileptic children.	Children with idiopathic temporal lobe epilepsy and IED in the left dominant hemisphere are at higher risk for CD than children with other types of IPE.	Papazian et al.	2003	Papazian O, Alfonso I, Garcia-Galarreta V. The effect of interictal epileptiform discharges on cognitive function in children with idiopathic epilepsy. Rev Neurol. 2003 Feb 1-15;36(3):282-4. Spanish.
Woodcock-Johnson	N/A	Seychelles	66 months	Creole (French-based)	Study to examine influence of social factors in relationship between	Gender significantly influenced the association between prenatal exposure and drawing and copying; however, the effects were not consistent.	Davidson et al.	2004	Davidson PW, Myers GJ, Shamlaye C, Cox C, Wilding GE. Prenatal exposure to

Test	Version	Country	Age range	Language	Purpose of Study	Results and conclusions	Author	Year	Complete Reference
					methylmercury (MeHg) exposure and MDI scores. Woodcock Johnson was used to test scholastic achievement.	Prenatal exposure interacted with one or more social or environmental covariates for general cognitive ability, overall language ability, and prearithmetic achievement. The effects were not consistent across either endpoints or covariate categories. Evidence of a small influence by social and environmental variables at 66 months is neither internally consistent nor consistent with earlier results. Overall, a consistent pattern of effect modification (EM) has not been observed, suggesting that the results may be due to chance.			methylmercury and child development: influence of social factors. Neurotoxicol Teratol. 2004 Jul-Aug;26(4):553-9.