ABSTRACT

Title of dissertation: A CONCURRENT VALIDATION STUDY OF THE MARYLAND DEVELOPMENTAL SCREEN

Abigail McNinch. Doctor of Philosophy, 2005

Dissertation directed by: Professor Paula Beckman, Department of Special Education

The purpose of this study was to examine the concurrent validity of a new screening instrument, the Maryland Developmental Screen (MDS), with the Bayley Scales of Infant Development, Second Edition (BSID-II). The MDS and BSID-II were concurrently administered at the University of Maryland (UMMS) Neonatal Intensive Care Unit (NICU) Follow Up Clinic to an age stratified sample of 81 infants and toddlers, born at 36 weeks gestation or younger, and whose corrected ages ranged from 18 days through 37 months of age. The sensitivity and the specificity rates were determined to be 88.2% and 90.1% respectively. The chance hit rate of 66.7% indicates that the overall hit rate of 90.1 was not due to chance alone. A relationship between the false negative and false positive outcomes in relation to the developmental domain of the child was determined. Evidence that the MDS has unique utility in identifying potential atypical development was also demonstrated. This study provides preliminary evidence that the MDS has validity in screening the development of infants and toddlers born prematurely. Further investigation of the instrument's validity using larger and more diverse sample sizes is needed.

A CONCURRENT VALIDATION STUDY OF THE MARYLAND DEVELOPMENTAL SCREEN

by

Abigail W. McNinch

Dissertation submitted to the Faculty of the Graduate School
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Committee:

Professor Paula Beckman, Chair
Associate Professor David Cooper
Associate Professor Brenda Jones-Harden
Assistant Professor Brenda Hussey-Gardner
Professor Joan Lieber

ACKNOWLEDGEMENT

I would like to thank my dissertation committee members, including Drs. Paula Beckman, David Cooper, Brenda Jones-Harden, Brenda Hussey-Gardner, and Joan Liber for helping design and implement the study and helping interpret the research the findings. Their feedback and input helped strengthen the quality of my work. Their comments were provided in a supportive and collaborative manner that made the process a valuable learning experience. To my dissertation chair and advisor, Dr. Paula Beckman, I would like to express my sincere appreciation for guiding me along the way through completion of my dissertation and degree. I am fortunate to have her as a mentor. Special gratitude goes to Dr. David Cooper who patiently provided me with a greater understanding of statistical procedures. I would also like to thank Joan Lieber and Brenda Jones-Harden for providing me insight and constructive feedback.

I would especially like to thank Dr. Brenda Hussey Gardner for being a supportive mentor. She has gone above and beyond to be a constant source of encouragement and has helped me handle many challenges along the way. She has provided me with valuable opportunities to work with infants born prematurely and their families. Without her help I would not have been able to meet my goals.

I would like to thank the wonderful families that took the time out of their busy schedules to participate in my research project. It was a pleasure to work with them. I also would like to thank the staff at the NICU Follow-Up clinic who helped with data collection. In particular, I would like to thank Dr. Christine Hess who was instrumental in my data collection, coding, and analysis. She also provided emotional support

throughout the process. I would also like to thank Kelly Tracey, whose assistance was invaluable.

Last, but not least, I would like to thank my family and friends. I am deeply grateful to my husband, Gene, for his love, encouragement, and proofreading throughout this challenging process. He has been a constant source of support, for which I am sincerely thankful. Finally, I would like to thank my daughters Hailey and Ainsley who patiently allowed their Mommy to get her work done.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

Introduction

In 2003, one in eight children in Maryland was born prematurely (March of Dimes, 2004). This high incidence is accompanied by an increased likelihood for these children to demonstrate developmental disabilities (Linden, Paroli, & Doran, 2000). Neonatal Intensive Care Unit (NICU) Follow-Up programs aim to identify children, born prematurely, who may be developmentally delayed; however, these programs lack valid screening measures to help do this effectively. The need for new valid screening tools is influenced by theory and driven by public policy (McLean, Bailey, & Wolery, 1996). The Maryland Developmental Screen (MDS) is a new screening tool, designed to screen the developmental status of children. Prior to this study, the MDS had not been subjected to validation procedures. The purpose of this first validation study was to determine the concurrent validity of the MDS when compared to the Bayley Scales of Infant Development, Second Edition (BSID-II) with a population of infants and toddlers who were born prematurely.

Although there has been no reduction in the incidence of prematurity, the mortality rate for children born prematurely has been reduced by half within the past 15 years (Hall, 2000). Despite the reduction in mortality, morbidity related to premature birth remains high. Approximately 11% of pregnancies result in premature births and these births are responsible for 70% of neonatal deaths and 50% of all neonatal disabilities. In the United States, an average of 6,040 children are born at a low birth weight (less than 2,500 grams) and 1,126 are born at a very low birth weight (less than 1,500 grams) each week (March of Dimes, 2004). Currently, the survival rate for infants

born at 23 weeks gestation is 20%, at 25 weeks gestation it is 65%, and between 26 and 33 weeks the survival rate ranges from 75% to 95% (Linden, et al., 2000). Of the 61.3% families that reported the ethnicity of their children who were born prematurely in Maryland between 2000 and 2002, the rates were as follows: 16.9% African American, 12.0% Native American, 11.7%, Hispanic, 10.8% Caucasian and 9.9% Asian (March of Dimes, 2004).

Although the survival rate has increased dramatically during the last two decades, a substantial body of research suggests that, despite advances in neonatal intensive care, infants who are born at a low birth weight (less than 2,500 grams) are at high risk for mental and physical disabilities (Berger, Holt-Turner, Cuppoli, Mass, & Hagerman, 1998; Linden et al., 2000; Resnick, Eyler, Nelson, Eitzman, & Bucciarelli, 1997). A lower birth weight (less than 1,500 grams) is related to a higher risk for developing future disabilities (McGrath, Sullivan, Lester, & Oh, 2000). Disabilities associated with premature birth vary and include cerebral palsy, language delays, and poor intellectual and neuromotor development, which often result in behavioral difficulties at school age (Nadeau, Boibin, Tessier, Lefebvre, & Robaey, 2001; McGrath, et al., 2000). According to Linden and her colleagues (2000), mild to severe disabilities occur with 66% of infants born between 23 and 25 weeks gestation, 60% of infants born between 26 and 29 weeks gestation, and 35% of infants between 30 and 33 weeks gestation.

Due to the risk of developmental delay demonstrated by infants born prematurely, it is important, especially within the first three years of life, to frequently screen and assess their developmental status. Screening, operationally defined as a “brief assessment designed to identify children who should receive more intensive diagnosis or assessment”

(Meisels & Provence, 1989, p. 58), is often used to examine large numbers of children in a cost- and time-efficient manner. Through the screening of a child's developmental status, appropriate anticipatory guidance may be provided to the family. The provision of anticipatory guidance involves informing caregivers about what a child is expected to do next as well as providing them with activities to help facilitate development. Subsequent assessment following the screening may also lead to referral to early intervention programs. The rationale for providing screening of infants born prematurely is grounded in developmental theory and is reflected in public policy.

Theoretical Frameworks

The theory which grounds developmental screening of infants and toddlers is reflected in many different disciplines, including special education, developmental psychology and behavioral pediatrics (McLean et al., 1996). These theories provide a basis for design, methods, and the subsequent validity of developmental screening instruments that are used. According to the maturationist theory, developmental screening and assessment of all children relies on an assumption that development occurs on a predictable continuum (Shonkoff & Meisels, 2000). The constructivist theory views children as active participants in the learning process (Shonkoff & Meisels, 2000). These two theoretical frameworks guided the prominent contributing perspectives of Als (Als, 1986; Als, Lester, Tronick, & Brazelton, 1982), Brazelton (1973), Gesell (1949) and Piaget (1952) in the assumption that child development is hierarchical and dependent on the environment.

Focusing on the neonatal period, Als and her colleagues' (1982) synactive model of development suggest that development occurs in a hierarchy of developing subsystems

of overall organization. These subsystems include motor (e.g., tone, movement, activity, posture), autonomic (e.g., skin color, tremors/startles, heart rate), states (e.g., sleepy/drowsy, awake/alert), attention/interaction (e.g., alertness) and self-regulatory (e.g., infants' ability to balance other subsystems) (LaRossa, 2002). Each of the five subsystems is dependent upon on one another and upon continuous interaction with the environment (LaRossa, 2000; McLean et al., 1996). LaRossa (2002), suggests that infants who are born prematurely typically have disorganized subsystems and are more dependent on the environment than infants who are full term and healthy. As a child grows, the subsystems mature, become more organized, and begin to promote one another (LaRossa, 2002). Brazelton's (1973) organismic view of infant development also suggests that infants' responses lead to and become a background for the subsequent levels of development. Like Als (1986), Brazelton's framework views an infant as an active contributor to development as demonstrated through his or her behavioral organization. (McLean et al., 1996).

Within the field of behavioral pediatrics, Arnold Gesell's (1949) theory of developmental schedules suggests that early development is maturational and that the environment determines the occasion, intensity and correlation of many behaviors (McLean, et al., 1996). The progression of development is inherently related to age-sequence development and this sequence of schedules is the most important indicator of a child's development. Many screening and assessment instruments, including the BSID-II (Bayley, 1993) and the Denver II (Frankenburg, Dodds, Archer, Shapiro, & Bresnick, 1992) have theoretical underpinnings based on Gesell's concept of developmental schedules (McLean et al., 1996).

Piaget (1952) theorizes that development occurs in a hierarchical series of stages. Each stage evolves from a preceding one, and no stage can be skipped (Sattler, 1992). An infant's sensorimotor organizational schemas, defined as interrelated memories, thoughts, and strategies that a child uses, change and mature with the assimilation and acquisition of new information (McLean et al., 1996). Screening instruments and assessments which draw upon a Piagetian approach are constructed according to age levels with item clusters that attempt to measure multiple aspects of development (McLean et al., 1996). The Piagetian approach to developmental screening and assessment is particularly reflected in the measurement of specific domains of cognitive development, including spatial concepts, object permanence, deductive and inductive logic, classification and decentration. Along with the theories of Als, Brazelton, and Gesell, Piaget's theory has influenced public policy that currently provides the delivery of early intervention services.

Policies and Mandates Required for Screening

The importance of developmental screening for infants and toddlers is reflected in federal legislation and is considered to be an integral part of the Child Find efforts. Part C of Public Law (P.L.) 99-457 of IDEA established Child Find for infants and toddlers in 1986. The Child Find effort requires that states identify all children eligible for education, health and social service programs (McLean et al., 1996). This mandate continues under the reauthorization of IDEA, P.L. 105-17, enacted in 1997. If a child is identified as eligible, then provisions, mandated by each state, are provided. In addition to the Child Find efforts, the most recent reauthorization of IDEA (2004) brought changes in Section 637 of Part C. This section states that every child involved with a substantiated case of

child abuse or neglect, as well as any child who is identified as affected by illegal substance abuse or withdrawal symptoms resulting from prenatal drug exposure, must also be developmentally screened to determine whether a referral for an evaluation under Part C is warranted.

In addition to Part C of IDEA, Early Head Start requires developmental screening within 45 days of a child's entry into the program (45 CFR 1304.20). The federal Early Periodic Screening, Diagnosis, and Treatment (EPSDT) program, based upon a model of preventative care for early detection of illness and developmental problems, also recommends routine developmental screening during well-child visits (Hess, Papas, & Black, 2004; Rosenbach & Gavin, 1998). Finally, the American Academy of Pediatrics (AAP) policy statement reflects the need for developmental screening, recommending that all infants and young children be screened for developmental delay at health supervision visits (Committee on Children with Disabilities, 2001). In addition to this policy statement, the AAP recommends that neurodevelopmental follow-up occur for all infants born prematurely (Committee on Children with Disabilities, 2001).

Importance of Validity

The effective implementation of public policies requiring the screening of infants and toddlers necessitates empirical evidence to assure that screening instruments used are valid. Validity refers to appropriateness of inferences that are made based on specific test results (Salvia and Ysseldyke, 2001.) One way in which an instrument's validity is demonstrated is through criterion validity. Criterion validity is an indicator of an instrument's accuracy in estimating performance on a widely accepted criterion measure (Salvia & Ysseldyke, 2001). Criterion validity research must include the following: (a)

an accurate description and rationale for use of the criterion measure, (b) a description of the sample and statistical analysis used, and (c) limits of generalizability of validity information (Salvia & Ysseldyke, 2001). It is also important to consider the consequential validity of screening tools (Messick, 1986). This concept is related to evidence of an instrument's actual and potential outcomes including value judgements, social implications, and political consequence (Humphries-Wadsworth, 1998).

The Problem

Neonatal Intensive Care Unit (NICU) follow-up programs are designed to provide ongoing screening and assessment of the growth and development of high-risk and moderate-risk infants discharged from the NICU. Once children are identified through such programs, early intervention services can be implemented. These programs do not typically replace traditional pediatric care; their purpose, rather, is specifically related to the identification of medical and developmental problems (Hussey-Gardner, Wachtel, & Viscardi, 1998). Many NICU follow-up programs (e.g., UMMS) use screening measures to determine whether or not a child's development may be delayed and therefore require in-depth evaluation. Existing screening instruments examine a child's developmental status through parent report, direct administration by a professional, or a combination of parent report and direct administration. The review of literature presented in Chapter II suggests that there are currently no published fully validated screening instruments that enable staff of NICU follow-up clinics to adequately measure the developmental status of infants born prematurely. At first glance, 12 screening instruments appeared to be capable of effectively screening for potential developmental delays of infants born prematurely. These instruments include:

- Ages and Stages Questionnaire (ASQ) (Squires, Bricker, & Potter, 1997),
- Healthy Steps "Quick Check Sheets" (Healthy Steps, 1996),
- Behavioral Assessment of Baby's Emotional and Social Style (BABES) (Finello, 1994),
- Child Development Inventories (CDI) (Ireton, 1992),
- Early Intervention Developmental Profile (EIDP) (Rogers, D'Eugenio, Brown, Donovan, & Lynch, 1981),
- Brigance Screen (Brigance, 1990),
- Brigance Infant Toddler Screen (BITS),
- Diagnostic Inventory for Screening Children (DISC) (Amdur, Mainland, & Parker, 1990),
- Battelle Developmental Inventory Screening Test (Newborg, Stock, Wnek, Guidubaldi, & Svinicki, 1998),
- Bayley Infant Neurodevelopmental Screen (BINS) (Alyward, 1995),
- Parent's Evaluation of Developmental Status (PEDS) (Ellsworth & Vandermeer, 1997),
- and the Denver Developmental Screen-II (Frankenburg, Dodds, Archer, Bresnick, Maschka, Edelman, & Shapiro, 1992).

However, none of these instruments best meet the needs for screening infants and toddlers born prematurely. Prior to this study, there was no established method to evaluate instruments designed to screen the development of infants and toddlers born prematurely. Based upon child development theory, and empirical and clinical evidence (Committee on Children with Disabilities, 2001; Glascoe et al, 1992; McLean et al.,

1996; Sonnader, 2000), the researcher identified the following criteria for NICU follow-up programs to most effectively screen children born prematurely (see Chapter II for further discussion):

1. developmental screening of all six developmental domains (namely cognitive, language, social-emotional, adaptive, gross motor, and fine motor),
2. use of minimal materials,
3. ease and speed of administration,
4. ability to test young infants with an adjusted age of less than term,
5. hands-on assessment,
6. parent involvement in the assessment process,
7. professional evaluation of the quality of the child's performance,
8. validity of screening.

Children should be screened in the six developmental domains so that the appropriate discipline (e.g., a speech and language therapist for a child with a language delay) may conduct a more in-depth assessment when needed. A screening instrument should be able to be administered with minimal materials and there should be ease and speed in administration so that clinic staff can easily screen many children in a NICU follow-up program's quick paced environment. Due to their degree of prematurity, children may attend an initial NICU follow-up program appointment prior to an adjusted age of full term (< 40 weeks gestation); a screening instrument should have the ability to assess development even at this point. Since hands-on assessment is a requirement of neonatology training programs, a screening instrument must have the capacity for administrators to observe and examine the children in addition to gaining information

from parent report. The requirement of parent involvement in the screening process is mandated in the federal legislation's Part C of P. L. 105-17 of IDEA (Part C – Individuals with Disabilities). In addition, professional evaluation of the quality of a child's performance should be provided. For example, if a child's gross motor development is age appropriate but he drags his left foot, the screening instrument should provide an opportunity for recording such a concern so that the child will optimally receive needed early intervention services. This is necessary because there is an atypical eligibility criterion for early intervention services and clinicians must have the tools to assist in the identification of this type of development. Finally, a screening instrument is only useful if it demonstrates validity (Committee on Children with Disabilities, 2001).

An in-depth analysis of the 12 instruments suggests that there is no measure that adequately meets each and every criterion and thereby meets the needs of a NICU follow-up program. Measures including the ASQ and the CDI fail to examine very young infants. An example of why this presents a problem is demonstrated with a child who was born 12 weeks early and is screened at three months of age. With an adjustment for prematurity, such a child should be tested at 40 weeks gestation, or at birth. Even at this early age, it is possible to gain pertinent developmental information regarding children; however, test items must be available to examine issues such as quality of tone and reflexes. Although the Denver II and other screening instruments do provide test items for children at birth, no information is obtained regarding the quality of performance observed. This issue also presents problems with parent report screening measures in which no professional observation of the child is conducted. Another problem is that results of studies also indicate that sensitivity and specificity rates of less than the

recommended 70% and 70-80% respectively (Sonnader 2000) have been found with screening measures including the Denver II, CDI and Brigance Screens (Byrne, Ashford, Johnson, Chang, & Strickland, 1992; Glascoe, 2002; Glascoe, Shoemaker et al., 1993).

The MDS, developed in 2001, attempts to meet the screening needs of children born prematurely (See Appendix A). This instrument's 161 items assess the six domains of cognition, language, social-emotional, adaptive, gross-motor and fine motor development of children whose adjusted ages range from 30 weeks post-conceptual age through 39 months of age. It does so in a manner that is quick and easy, uses minimal materials, and uses information gathered by both hands-on assessment and parent involvement in the assessment process. Although currently in use in one NICU follow-up program, prior to this study the MDS had not yet undergone validity procedures.

Purpose and Objectives

An examination of currently available screening instruments suggests that there is no published tool that adequately assesses the developmental status of infants born prematurely. Although one instrument, the MDS, appears to meet this unmet need, it had yet to be subjected to rigorous validation procedures. The purpose of this study was to determine the concurrent validity of the MDS when compared to the "gold standard" (Sattler, 1992, p. 321) of developmental assessments, the Bayley Scales of Infant Development, Second Edition (BSID-II). To address this issue the following research questions were asked:

1. What is the sensitivity of the MDS when concurrently administered with the BSID-II?
2. What is the specificity of the MDS when concurrently administered with the BSID-II?
3. What is the overall hit rate and chance hit rate of the MDS when compared to the BSID-II?

CHAPTER II

Review of Literature

This chapter presents a review of the literature in three important areas and then describes the MDS and its application. This chapter will provide an understanding of the developmental and medical outcomes of children born prematurely. The function of Neonatal Intensive Care (NICU) follow up programs that often monitor the progress of these children is also described. Developmental screening of infants and toddlers, including a critical examination of published screening tools, is presented. Finally, a description of the Maryland Developmental Screen (MDS) and the instrument's application will be provided.

Outcomes of Children Born Prematurely

While the survival rate for children born prematurely has dramatically increased within the last 30 years, longitudinal studies suggest that those children who are born at a very low birth weight (less than 1,500 grams) and extremely low birth weight (less than 1,000 grams) are more prone to developmental delay (McGrath et al., 2000; Nadeau et al., 2001; Perlman, 2001). Medical complications due to low birth weight may impact later development. Complications may include underdevelopment of the lungs, digestive system and nervous system; more specifically, these infants are at risk for necrotizing enterocolitis, intraventicular hemorrhage, respiratory distress syndrome, broncopulmonary dysplasia, patent ductus arteriosis, and periventricular leukomalacia (Bernbaum & Batshaw, 1997; Subramanian, Yoon, & Troal, 2002). Neurodevelopmental, intellectual, motor, and language outcome have been demonstrated to be negatively

associated with low birth weight (McGrath et al., 2000; Nadeau et al., 2001; Singer, Siegal, Lewis, Hawkins, Yamashita, & Baley, 2001).

There has been a growing body of longitudinal research related to the developmental outcomes of infants born prematurely (Hack, Taylor, Klein, Eiben, Schatschneider & Mercuric-Minich, 1994; McGrath et al., 2000; Nadeau et al., 2001; Singer et al., 20001; Vorh, 2000). Perlman (2001) found that large number of infants born prematurely exhibit neurobehavioral problems, even in the absence of cerebral palsy. Approximately 30% to 50% of children born prematurely demonstrate below average academic achievement, 20% to 30% are affected by attention deficit/hyperactivity disorder, and approximately 25% to 30% demonstrate psychiatric disorders during adolescence (Perlman, 2001). Perlman (2001) attributes these neurobehavioral problems to an inherent vulnerability of a prematurely developed brain during a critical period of development, the multiple clinical problems which are specific to prematurity, and the stressful environmental conditions in which a premature infant is placed.

Similarly, researchers suggest that behavioral problems at school age may be related to the influence of prematurity on neuromotor functioning (Nadeau et al., 2001; McGrath et al., 2000). Neurodevelopmental deficits demonstrated by children born prematurely are often manifested in the form of behavioral problems and academic difficulties. Nadeau and others' (2001) research with 61 very low birth weight (VLBW) and 44 normal birth weight children found that elementary school children, between five and nine years of age who were born at VLBW were at significantly increased risk for neurological problems (33%) compared to children born at a normal birth weight (3%).

Parents and teachers observed that VLBW children have more internalized and externalized problems than their peers who were born at a normal weight. The internalized problems included more social withdrawal and sadness. External problems cited were hyperactivity, inattentiveness, and aggressiveness. Peers of the children with VLBW also evaluated them to be more sensitive and hyperactive than children born at a normal birth weight. Results of this study provide evidence that preterm birth is associated with intellectual and neuromotor delays and that these deficits account for the predictive relationship between premature birth and behavioral problems (Nadeau et al., 2001).

McGrath and others (2000) also studied neurological functioning with a sample (N = 188) of children who were born at full term, healthy preterm, clinically ill, and neurologically compromised. The sample of children was assessed at 18 and 30 months and at four and eight years of age. The methodological procedures conducted involved repeated neurological categorization and developmental assessment of cognition, academic performance, socioeconomic status (SES), and medical status. Similar to the work of others (Perlman, 2000; Nadeau et al., 2001), findings of this study suggest that the change in neurologic classification over time varies as a function of neonatal morbidity, and the changes identified also affect cognitive and school achievement outcomes. McGrath and others (2000) indicate that this study extends the research by identifying neurological examination throughout childhood as a marker for long-term outcomes of NICU survivors.

Longitudinal research of infants born prematurely conducted by Singer and others (2001) has focused on language development. These researchers found that 20%-40% of

VLBW survivors are considered language delayed or impaired as toddlers and young children. The researchers (Singer et al., 2001) conducted a prospective study of infants born at VLBW with and without bronchopulmonary dysplasia (BPD) and of infants born at full term to examine speech and language development and specific language impairments at three years of age. Findings suggest that infants born at VLBW with a history of BPD have lower receptive and language skills than infants born at a VLBW with no history of BPD and than those born at full term. A surprising result of this study was that the presence of patent ductus arteriosis (PDA) was the best predictor of language deficits, and when the occurrence of PDA and BPD were combined, there were differentially lower language scores (Singer et al., 2001).

Prematurity and Socioeconomic Status (SES)

The development of a child born prematurely is likely to be complicated, particularly as that child matures, by environmental risk factors such as low SES, low maternal age, and teenage parenting factors (Hess, Papas, & Black, 2004; Leonard et al., 2001). Researchers have examined longitudinal developmental outcomes of infants born prematurely in relation to how the environment impacts development (Hack et al., 1994; Singer et al., 2001; Weisglas-Kuperus, Baerts, Smrkovsky, & Sauer, 1993). Singer et al. (2001) indicate that neurological complications, low socioeconomic status, and minority race (not specified) were significant predictors of language delay. They caution, however, that it is important to consider both medical and SES factors when evaluating the risk of infants born at VLBW for poorer speech language outcomes (Singer, et al., 2001). Hack and his colleagues (1994) similarly suggest that although developmental outcomes appear to be more closely associated with neonatal complications than social

disadvantage, both should be considered. The research conducted by Weisglas-Kuperus and her colleagues (1993) also examined the relationship of sociodemographic variables, social stimulation and cognitive development on infants ages 1 through 3.6 years who were born at a VLBW. Neurological scores obtained through the Kaufman Assessment Battery (Melchers & Preuss, 1991) predicted infant development in the first year of life; from two years of age, however, a combination of biological and home environment factors influenced child development. Environmental factors become more significant as a child matures, but biological factors, such as prematurity, may be more important in predicting the first two years of a child's development (Hess et al., 2004).

NICU Follow-Up

Findings within the literature suggest that when children who were born prematurely are discharged from the NICU they remain at risk for future developmental disabilities (Montgomery, 1999; Bull, Bryson, Schreiner, & Lemons, 1986). This risk necessitates systematic monitoring, follow-up, and early intervention services. The intent of a NICU follow-up clinic is to provide "ongoing evaluation of growth and development of high risk and moderate risk infants discharged from the NICU" (Hussey-Gardner, 1995, p. 33). Rather than replace routine pediatric care, NICU follow-up clinics should provide examinations and testing designed to recognize early developmental and medical problems (Hussey-Gardner, 1995). NICU follow-up programs are designed to provide a coordinated effort of identification, evaluation, and service delivery to infants. A recommended best practice of care for these infants involves the collaboration of neonatal intensive care unit staff, early intervention providers, and outpatient NICU follow-up

staff in order to ensure appropriate referral and expedient delivery of early intervention services (Hussey-Gardner, McNinch, Anastasi, & Miller, 2002).

Developmental Screening

Developmental screening is defined as "a brief assessment designed to identify children who should receive more intensive diagnosis or assessment" (Meisels & Provence, 1989, p. 58). Screening may result in the assessment of a child's developmental status and subsequently lead to referral to early intervention programs. Screening is based upon a framework of typical development in which a child's achievements are placed upon a continuum of normal accomplishments (Shonkoff & Meisels, 2000). Assessment differs from screening because it involves a more in-depth collaborative process of ongoing, systematic observations and analysis (Greenspan & Meisels, 1996).

Age Adjustment. An adjustment of age to account for a child's prematurity is often calculated prior to conducting developmental screening and/or assessment. This adjustment allows for a more accurate comparison of the developmental status of children born prematurely with children who are born at full term (Brenbaum & Batshaw, 1997). When using the child's chronological age rather than an adjusted age, studies have found that achievement of developmental milestones occurs significantly later than with children who are born at full term (Matilainen, 1987). There is, however, disagreement and inconclusive research regarding at which age this adjustment should be discontinued (Brenbaum & Batshaw, 1997; Rickards, Kitchen, Doyle, & Kelly, 1989; Wilson & Michaeleen, 2004). According to Bernbaum and Batshaw (1997), the classic approach to age adjustment is to continue do so until the child is two years of age, thereby presuming

that by this age he or she has caught up with children of the same age who were born at full term. However, as infants mature, a complete age adjustment may lead to an overestimation of a child's development status (Wilson & Michaeleen, 2004). Because there is no empirical consensus as to the age at which the adjustment should stop, programs often set their own criteria. For example, at the University of Maryland's NICU Follow-Up Clinic, the age when the adjustment is stopped is when the child's adjusted age is one year old.

The process of screening infants and young children is often the first experience that a family has with early intervention and frequently also serves as a therapeutic experience in itself (McLean, et al., 1996). According to Harris and Daniels (2001), most pediatric developmental screening tests are designed with the intent of differentiating children who are suspected of having a developmental concern from those who appear to be developing normally. A screening instrument enables evaluators to cast a wide net to select children who may need closer monitoring as well as serving as a formal indicator of present developmental status and need for timely early intervention services (Harris & Daniels, 2001; Leonard, Piecuch, & Cooper, 2001; Sonnader, 2000).

Researchers recommend that developmental screening be in the form of a brief evaluation, intended to identify children with suspected problems who are in need of more complete diagnostic assessment (Doig, Macia, Conway, Craver, & Ingram, 1999). Criteria for evaluating screening tests should include an examination of the variables of acceptability, simplicity, cost appropriateness, and reliability and validity (Harris & Daniels, 2001). The use of a screening tool must be acceptable to the family whose child is being screened, to the professionals who administer it, and to the community in which

it is used. Additionally, it should be relatively simple in terms of training and administration.

Doig and colleagues (1999) stress that repeated screening is necessary because both the biological and environmental risk factors that affect development can change over time. A single test at one point in time provides a snapshot of a child's developmental process; periodic screening is necessary to detect emerging delays as a child grows (Committee on Children with Disabilities, 2001). Screening in the first year of life, particularly with an infant born prematurely, can be problematic. Complications may include transient muscle tone differences that may present as a delay and often resolve by the first year. Furthermore, children born at a low birth weight have increased medical risk factors such as asphyxia and maternal substance abuse, as well as increased social risk factors including low SES, low maternal age, and teenage parenting (Leonard et al., 2001).

Subspecialty pediatric clinics, as well as general pediatrician offices, utilize screening instruments and developmental assessments to monitor the development of children. When pediatricians and other health care providers use only their clinical judgment rather than formal screening and assessment procedures, estimates of a child's development may be less accurate than when a more objective instrument is used (Committee on Children with Disabilities, 2001). Limited time and low reimbursement rates, however, hamper the use of developmental screening and assessments (Doig et al., 1999). Use of formal screening in pediatric practice is also limited by lack of consensus on what is suitable for screening general populations, what is easily integrated into the practice protocol, and what is cost efficient (Dobrez, Sasso, Holl, Shalowitz, Leon, &

Budetti 2001). Screening instruments are designed to obtain information on a child's developmental status using questionnaires provided to the parents, through administration of test items by professionals, and through recording observations of children.

Sensitivity and Specificity

A screening instrument's level of sensitivity and specificity should be examined when considering use. An instrument's sensitivity is related to how accurately children with developmental delays are identified and the specificity refers to the accuracy in identifying those children without disabilities (Sonnader, 2000). The evaluation of an instrument's levels of sensitivity and specificity is important in determining cutoff scores for making referrals for further assessment, which may consequently lead to the acquisition of early intervention services (McLean et al., 1996). For example, in Maryland, one way that a child is determined to be eligible for early intervention services is if he or she demonstrates a 25 % delay in one or more developmental areas. The levels of sensitivity and specificity of a particular screening instrument are, therefore, vital in making decisions regarding which children should receive in-depth assessment to determine whether they are deemed eligible for services. According to Sonnader (2000), a level of 70% for sensitivity and a level of 70-80% for specificity are acceptable percentages for developmental screening instruments.

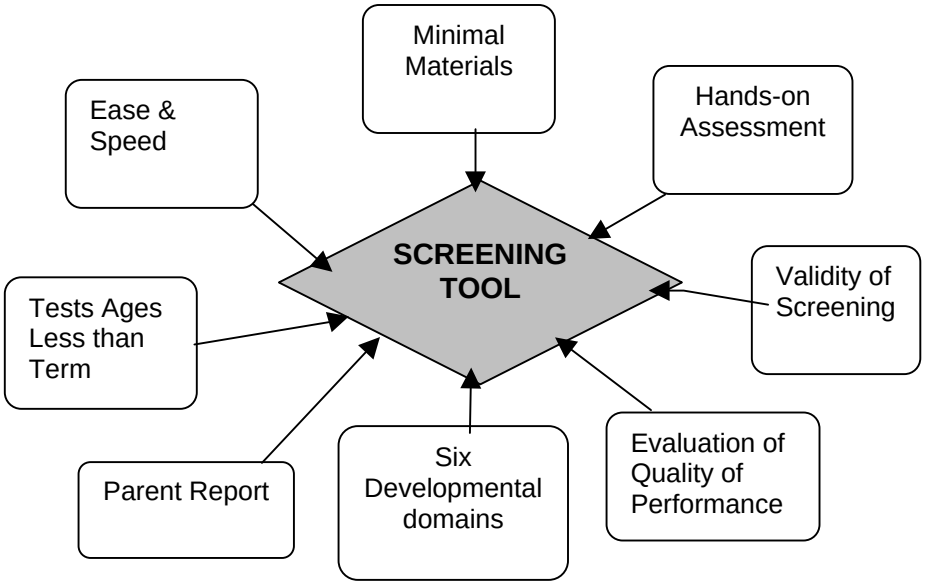
Screening Instruments

Accurate developmental screening of infants and toddlers contributes to parental well-being and assists in appropriately allocating limited diagnostic and health care services (Glascoe, Byrne, Ashford, Johnson, Chang, & Strickland, 1992). Early childhood screening also serves as a reminder to professionals to observe a child's development and provides an efficient method to record clinical observations (Committee on Children with Disabilities, 2001). Most importantly, however, screening ensures that children who are in need of early intervention due to the presence of a developmental delay are identified as early as possible (Committee on Children with Disabilities, 2001; Glascoe et al., 1992; McLean et al., 1996; Sonnader, 2000).

Although at first glance there may appear to be a plethora of screening instruments widely available to determine the developmental status of infants and young children born prematurely, a closer look at the literature yields few viable instruments. The literature was reviewed to determine whether there were any existing screening tools that adequately meet the needs of a NICU follow-up program. The criteria upon which each tool was evaluated presented in Figure 1 included: (a) developmental screening of all six developmental domains (namely cognitive, language, social-emotional, adaptive, gross motor, and fine motor), (b) use of minimal materials, (c) ease and speed in administration, (d) ability to test young infants with an adjusted age of less than term, (e) hands-on assessment, (f) parent involvement in the assessment process, (g) professional evaluation of the quality of a child's performance, and (h) validity of the test. Furthermore, the technical adequacy of instruments considered for the use with infants born prematurely was explored.

Figure 1

Criteria for Developmental Screening of Infants and Toddlers Born Prematurely



Professionals have accomplished the screening of infants and young children through direct administration of test items, parent report of child development, and a combination of administration and parent report. Instruments attempting to screen only a child's global development were omitted from this review. A rationale for the decision to omit instruments that approach developmental screening through examination of global development, such as the Vineland Social Maturity (Doll, 1953), is provided by Katoff and Reuter (1980). These authors suggest that the separation of items into developmental domains has value that is prescriptive, diagnostic and predictive. Whereas a typically developing infant generally exhibits uniformity across all domains of growth, an infant who is delayed may demonstrate unique patterns and inconsistencies. Screening instruments that provide only global scores or fail to separate development into domains consequently fail to yield adequate information to indicate the direction of further

assessment and intervention. Furthermore, these are instruments not particularly useful in providing feedback to parents (Katoff & Reuter, 1980). Also omitted were instruments that screen one domain only. These instruments were not included as they do not provide information across domains. Domain specific instruments omitted include the Cognitive Adaptive Test and Clinical Linguistic and Auditory Milestone Scales (CAT/CLAMS) (Capute & Accardo, 1996) and the Peabody Developmental Motor Scales (Folio & Fewell, 2000).

Screening measures that exclusively examine neonatal neuromotor function were additionally excluded from this review. Instruments examining this newborn period are not typically designed to assess the development of older infants and toddlers. These instruments, including the Brazelton Neonatal Behavioral Assessment Scale (Brazelton, 1973), the Carey Temperament Scales (Carey, 2002), and the Harris Infant Neuromotor Test (Harris & Daniels, 2001) focus on normal qualitative behavioral variations and are not designed to be used to screens for early identification of developmental or behavioral abnormalities (Carey, 2002). Rather, the general purpose of such neonatal measures is to demonstrate an infant's "capacities for using his or her inner organization to experience, integrate, and profit developmentally from developmental stimulation" (Brazelton, 1994, p. 289).

Professionals have accomplished the screening of infants and young children through direct administration of test items, parent report of child development, and a combination of administration and parent report. Table 1 provides an overview of screening tools included in this review. An overview of studies examined, including statistical indices, is provided in Table 2.

Table 1

Screening Tools

Tool	Authors	Ages	Method	Areas Assessed
Ages & Stages Questionnaire (ASQ)	Bricker & Squires (1997)	4 months to 5 years	Parent report	Communication, gross motor, fine motor, problem solving, personal-social
Behavioral Assessment of Baby's Emotional & Social Style (BABES)	Finello (1994)	High-risk infants	Parent report	Temperament, ability to self-soothe, regulatory processes
Battelle Developmental Screening Test	Newborg, Stock, Wnek, Guidubaldi, & Svinicki (1988)	Birth to 8 years	Direct & parent report	Gross motor, fine motor, psychosocial, adaptive expressive language, receptive language, cognitive
Bayley Infant Neurodevelopmental Screen (BINS)	Alyward (1995)	Birth to 42 months	Direct & parent report	Neurological processes, neurodevelopmental skills, developmental accomplishments
Brigance Infant Toddler Screen (BITS)	Brigance & Glascoe (2002)	Birth to 2 years	Direct	Fine motor, receptive language, expressive language, gross motor, self-help, social-emotional
Brigance	Brigance (1990)	21 to 36 months	Direct	Fine motor, receptive language, expressive language, gross motor, self-help, social-emotional
Child Developmental Inventories (CDI, formerly Minnesota Developmental Inventories)	Ireton (1992)	15 months to 6 years	Parent report	Social, self-help, motor, language, letter and number skills, presence of symptoms and behavior problems
Denver II	Frankenburg et al, (1992)	Birth to 6 years	Direct & parent report	Gross motor, fine motor/adaptive, language, personal-social

Table 1: Screening Tools continued

Tool	Authors	Ages	Method	Areas Assessed
Diagnostic Inventory for Screening Children, 3rd Edition (DISC)	Amdur, Mainland, & Parker (1990)	Birth to 5 years	Direct	Expressive language, receptive language, gross motor, fine motor, psychosocial, self-help
Early Intervention Developmental Profile (EIDP)	Rogers, D'Eugenio, Brown, Donovan, & Lynch (1981)	Birth to 36 months	Direct	Perception/fine motor, cognition, language, social-emotional, self-care, gross motor
Parents Evaluation of Developmental Status (PEDS)	Ellsworth & Vandermeer Press, Ltd. (1997)	Birth to 8 years	Parent report	Cognition, expressive language and articulation, receptive language, fine motor, gross motor, behavior, social self-help, school skills
Quick Check Sheets	Healthy Steps (1996)	Birth to 3 years	Parent report	Developmental areas, parent-child interactions, parenting concerns

Table 2

Overview of Selected Studies

Tool	Study Authors	Comparison Criteria	N	Indices
ASQ	Squires, Bricker, & Potter, 1997	standardized measures	7,000	Specificity: high across questionnaire intervals Sensitivity: lower, varied across questionnaire intervals Concurrent validity: 85% (range of 76% to 91%)
BINS	Leonard, Piecuch, & Cooper, 2001	BSID-II (MDI & PDI)	133	Association: MDI r = .4 PDI r = .35 Predictive validity: 67-76%
BINS	Hess, Papas, & Black, 2004	BSID-II (MDI & PDI)	106	6 & 13 month BINS scores used to predict BSID-II at 24 months: low sensitivity values, high specificity values
CDI	Shoemaker, Saylor, & Erickson, 1993	BSID-II (MDI)	280	Concurrent validity: .39 to .53 for all scales Specificity: 92% Sensitivity: 56%
CDI	Kopparthi, McDermott, Sheftel, Lenke, Getz, & Frey, 1991	BSID-II (MDI & PDI)	101	Correlation: strong correlation by domain
CDI	Diog, Macias, Saylor, Craver, & Ingram, 1999	CAT/CLAMS BSID-II	73	Correlation: CAT/CLAMS r = .87 BSID-II r = .86 Specificity: 94-96% Sensitivity: 80-100%
CDI	Montgomery, 1999	CAT/CLAMS Slosson Intelligence Test	76	Specificity: 87% Sensitivity: 73%
CDI	Saylor & Brant, 1986	BSID-II	115	Correlation: .75 - .91

Table 2: Overview of Selected Studies continued

Tool	Study Authors	Comparison Criteria	N	Indices
Denver II	Glascoe, Byrne, Ashford, Johnson, Chang, & Strickland, 1992	BSID-II Kaufman Assessment Battery for Children, Stanford-Binet 4th ed., Vineland Adaptive Behavior Scale	104	Specificity: 43% Sensitivity: 83%
DISC	Schwarting, 1998	other screening tests & criterion measures	*	Validity: Good content and face Reliability: Adequate
BITS	Glascoe, 2002	BSID –II, REEL, Preschool Language Scale, Rosetti, Vineland, Alberta Infant Motor Scale	408	Specificity: 86% Sensitivity: 77%
Battelle	Glascoe & Byrne, 1993	intelligence, adaptive, language, and achievement measures	104	Specificity: 73% Sensitivity: 75%
Battelle	Ikle & Wittmer, 1995	Vineland BSID-II Denver II		Validity: High Evidence of over-referral compared to Denver II

* Unavailable

Parent Report

The design of screening instruments that utilize parent report to determine the developmental status of infants and young children varies from checklists of questions asked by professionals to questionnaires that are completed by the parents. Parent report instruments are traditionally attributed with good psychometric properties and, because they are quick to administer, are often employed in pediatric offices (Committee on Children with Disabilities, 2001). Published screening instruments that elicit developmental information via parent report include the Ages and Stages Questionnaire (ASQ), the Healthy Steps "Quick Check Sheets", the Behavioral Assessment of Baby's Emotional and Social Style (BABES), and the Child Development Inventories (CDI).

Questionnaires, such as the ASQ, provide directions for parents to indicate their child's skills. Although questionnaires are often cost effective and take little time from the professional, concerns related to length of the questionnaire and accuracy of parental response should be considered. Researchers in one study suggest that questionnaires that are provided to caretakers yield a low return, as one-third of their sample was compliant (Doig et al., 1999).

Ages and Stages Questionnaire (ASQ). The ASQ is a screening instrument that focuses on children from four months through five years of age. Drawings and directions are provided to parents so that they can indicate their child's skills. Parents complete a 30-item questionnaire given at 11 different intervals based on their child's chronological age (4, 6, 8, 12, 16, 18, 20, 24, 30, 36, and 48 months of age). Each questionnaire probes parents on information related to the five developmental areas of gross motor, fine motor, communication, cognition, and personal-social development (McLean et al., 1997).

Included are questions geared to assess neurological reflexes (e.g., reflexes and tone), neurodevelopmental skills (e.g., movement and symmetry), and developmental accomplishments (e.g., object permanence, imitation, and language) (Dobrez et al., 2001). Items are scored by a check of "yes", "sometimes", or "not yet", and in turn are given a numeric value. Within each domain the numeric score is compared with cutoff scores found within the manual and if the child's score is below the cut-off, more in-depth assessment is recommended.

Squires, Bricker and Potter's (1997) validity research suggests that specificity of the ASQ is high across questionnaire intervals. Although advantages of utilizing the ASQ include cost-effectiveness, parental involvement and flexibility in administration procedures, it is not an appropriate instrument for use in a NICU follow-up program. It is less than optimal for several reasons. Most importantly, the instrument begins screening children at four months of age; a NICU Follow-Up clinic often sees children whose developmental age, once an adjustment for prematurity has been made, is still less than full term. For example, a six month old child born 28 weeks gestation (12 weeks premature) has an adjusted age of three months old and consequently would not be able to be appropriately screened until he or she was seven months of age. Additionally, as one of a NICU follow-up program's functions is to provide hands-on fellowship training for physicians, a neonatologist must monitor the children's development; therefore a questionnaire based screening instrument is not desirable.

Behavioral Assessment of Baby's Emotional and Social Style (BABES). The BABES is a short checklist of social and emotional behaviors of children. Designed for a high-risk infant population, this instrument focuses on temperament, ability to self-

soothe, and regulatory processes (Dobrez et al., 2001; Leonard, et al., 2001). Examples of items include transferring of blocks, looking for fallen items, types of vocalizations, and pre-walking progression. Neurological intactness in relation to active and passive tone in the upper and lower extremities and quality of movement is also examined. The BABES has been validated on high-risk infant populations as wall as in normal test construction. Concurrent validity has been conducted with two other instruments utilizing a population of infants born at term and preterm, and at high risk for developmental delay (Macias, Saylor, Greer, Charles, Bell, & Katikaneni, 1998).

The BABES is not the most optimal screening instrument for use in a NICU follow-up clinic as this measure focuses more on identifying social and emotional problems rather than determining whether there is a need for in-depth developmental assessment.

Child Development Inventories (CDI). Formerly the Minnesota Development Inventories (MDI), the CDI was designed to provide "systematic ways of obtaining in-depth developmental information from parents" (Ireton, 1992, p. 1). The author of the CDI suggests that this measure enables parents to become collaborators rather than passive observers in the assessment process. (Ireton, 1992). The CDI utilizes three separate instruments with 60 yes/no descriptions (Glascoe, 2002). This parent questionnaire assesses the eight developmental domains of social, self-help, gross motor, fine motor, expressive language, language comprehension, letters and numbers. It addition, a General Developmental Scale is included. The CDI is designed to developmentally screen children between the ages of 15 months and 6 years. Similar to other screening tools, the age range of children served in a NICU follow-up program

extends below the CDI's scope. An additional concern with this instrument is that various researchers (Diog et al., 1999; Shoemaker, 1993) demonstrate a large discrepancy of levels of sensitivity among various studies.

Additional Instruments. In addition to the parent report screening tools described in detail, others exist but were excluded from use without an in-depth exploration because they failed to meet the criteria of use in one NICU follow-up program. The Healthy Steps "Quick Check Sheets" and the Parents' Evaluation of Developmental Status (PEDS) are two such instruments. The Healthy Steps "Quick Check Sheets" provides professionals with an overview of what to discuss at visits, a checklist of developmental areas to observe, a checklist of parent-child interactions to observe, and questions to help parents discuss parenting concerns (Dobrez et al., 2001). This instrument provides professionals with a tool to aid in identification of potential concern but does not provide ample information to be utilized as a NICU follow-up program's screening measure. The PEDS is standardized to elicit parent concerns related to their child's development. This screening instrument is appropriate for children from birth through the age of eight. Ten questions are used to elicit information and include one on cognition and eight related to the developmental domains of expressive language and articulation, receptive language, fine motor, gross motor, behavior, social self-help, and school skills. Again, the brevity of this tool deemed it inappropriate for use in a NICU follow-up program.

Direct Administration

Some screening instruments are directly administered by professionals or by trained paraprofessionals. Examples of direct administration include observation of a child attempting a given task, direct elicitation from specific directions, and manipulation

of the child's environment. These instruments do not necessitate nor encourage parental involvement in the screening process. Examples of instruments utilizing direct administration only include the Early Intervention Developmental Profile (EIDP), the Brigance Screens, and the Diagnostic Inventory for Screening Children (DISC).

Early Intervention Developmental Profile (EIDP). Considered a screening tool by some (Ikle et al., 1995), the EIDP is a tool used to examine child development, focusing on children ages birth through 36 months. Designed to bridge the gap between assessment and program implementation, the EIDP is not a predictor of a child's future capabilities; rather it is an aid in predicting what type of skills a child is expected to develop next (Rogers et al., 1981). Six scales include information on perception/fine motor, cognition, language, social-emotional, self-care, and gross motor development. In addition to the lack of opportunity for parent reporting of a child's abilities, this instrument is not recommended for use with infants born prematurely because it does not assess children born at less than 40 weeks gestation.

Brigance Screens. Developed in the early 1980's, the Brigance Inventory of Early Childhood Development is an instrument which assess the six developmental areas of fine motor, receptive language, expressive language, gross motor, self-help, and social emotional skill (Glascoe, 2002) of children ages 21 through 36 months. More recently a new tool, the Brigance Infant and Toddler Screen (BITS), has been developed and extends chronologically downwards to screen children from birth through two years. There are two versions of the BITS, direct elicitation/observation and parent-interview/self-report. Quicker than instruments such as the Denver II, the BITS can be administered in 10 to 15 minutes. The Brigance screening instruments were not

considered for use in a NICU follow-up program as a combination of the two Brigance screening instruments would be required to meet the assessment needs of the clinic, does not assess children whose chronological age is less than full term, and, similar to the ASQ, omits examination of the cognitive domain.

Diagnostic Inventory for Screening Children (DISC). Developed in Canada in the late 1970's, the DISC examines developmental status of children from the ages of birth through five years. This instrument examines eight areas including fine motor, receptive language, expressive language, gross motor, auditory attention and memory, visual attention and memory, self-help and social development. Although this instrument is referred to as a screening measure, Schwarting (1998) suggests that it is better described as a useful instrument when administered between a general screen and full developmental evaluation. Several researchers agree that the materials used are not appropriate, such as the use of buttons and items with small detachable pieces. Furthermore, the test kit does not provide all items, such as a mirror, utensils and food, and stairs (Schwarting, 1998; Watson & Henington, 1998).

Despite Schwarting's (1998) reporting adequate content and face validity, other reviewers of the DISC (Watson & Henington, 1998) cite various technical problems. The small standardization sample used a Canadian population (French, English, other) and, therefore, is not reflective of an American population. More of a concern is the omission of children with disabilities within this sample, of special importance as the DISC was designed to identify such children. Finally, there is no data regarding concurrent, or predictive validity (Watson & Henington, 1998).

Combination of Direct Administration and Parent Report

In 1975 P. L. 94-142 of the Individuals with Disabilities Education Act (IDEA) included the provision that parents be part of the process for evaluating their child and that parents have the authority to challenge the accuracy of the evaluation, program, or placement of their child (Turnbull, Turnbull, & Wheat, 1982). Recent recommendations for best practice in early intervention views families as equal members who should join together with staff and take part in the early intervention system, including all aspects of their child's care and at all levels of decision making (Mahoney et al., 1999). The screening process is a vital component of early intervention in which parents must be viewed as members participating in the team rather than as clients of the professionals. Screening instruments that combine direct administration and parent report enable parents to be participants of this process. A combination of professional observation and parental input provides the most accurate picture of a child's development. Screening measures which combine direct administration and parent report include the Battelle Developmental Inventory Screening Test (Battelle), Bayley Infant Neurodevelopmental Screen (BINS), and the Denver Developmental Screen-II.

Battelle Developmental Inventory Screening Test (Battelle). Developed in 1984, the Battelle was designed to screen children, ages birth through eight years in the domains of cognition, receptive language, expressive language, gross motor, fine motor, psychosocial, and adaptive development (Ikle & Wittmer, 1995). This screening instrument is designed to be administered by professionals or trained paraprofessionals and takes 10 to 20 minutes to complete. According to Ikle and Wittmer (1995) the validity of the Battelle is reported as high when compared with other instruments

including the Vineland and BSID-II. Glascoe and others (1992) have also determined the use of the Battelle to be well supported through validity studies. Ikle & Wittmer (1995), however, express concerns related to the specificity of this screening instrument as there is evidence of over-referral when compared to the Denver-II.

Although the Battelle included a diagnostic evaluation, it is not recommended for children ages six months or younger, and therefore would not be a tool appropriate for use with a NICU follow-up program's population (Ikle & Wittmer, 1995). Furthermore, according to Ikle and Wittmer (1995) this screening test provides few items at each age level (the total number of items is 85) when compared to other measures.

Bayley Infant Neurodevelopmental Screen (BINS). The BINS is a screening instrument developed from items within the BSID-II. An administrator directly elicits 11 to 13 items per three to six month age range within approximately a ten-minute time frame. Assessing the development of children from the ages of three months through 24 months, each age range examines the child's neurological processes, neurodevelopmental skills and developmental accomplishments (Dobrez et al., 2001; Krug, 1998). Leonard and others (2001) examined the appropriateness of the BINS as a screening technique for children born prematurely at a low birth weight. The researchers suggest that the BINS is a satisfactory screener for infants born prematurely and at low birth weight when used in conjunction with known biologic and social risk factors (Leonard et al., 2001).

Other researchers found the BINS to demonstrate low sensitivity but high specificity when predicting BSID-II results (Hess, Papas, & Black, 2004). Their sample included low income African American infants who were classified as high risk and were administered the BINS at 6 and16 months and the BSID-II at 24 months. Hess and her

colleagues (2004) suggest that the BINS is an appropriate measure of developmental risk for children who are at biological risk (e.g., premature infants) because it may capture skill that continue to be problematic over time. They argue that the BINS is not, however, an ideal measure to predict development of infants who are environmentally at risk with no biological risk factors because infants who are in high risk environments have heightened risk for developmental delays after infancy (Hess et al., 2004).

The BINS strengths are related to a good normative sample, good organization of the areas assessed and the scoring procedures (Benish, 1998; Krug, 1998). It is recommended that administrators possess a graduate level knowledge of infant development and handling. This high level of knowledge needed may result in limited use of the BINS in educational and non-clinical settings (Benish, 1998). Krug (1998) further cautions that despite the BINS' strength of identifying low- to moderate-risk children who are developing appropriately, there is risk that the instrument also may incorrectly categorize children as high to moderate-risk infants. These false positive identifications may lead children to be subjected to lengthy, unnecessary testing. The main justification, however, for not using the BINS with a NICU follow-up program population, however, is that this measure does not assess the development of children less than three months of age.

The Denver II Developmental Screening Test (Denver II). The Denver II is the most widely used screening throughout the world (Dobrez et al., 2001). This screening instrument is administered to children between birth and six years of age in order to determine performance on age-appropriate tasks (Frankenburg et al., 1992). Similar to the MDS, the 125 item Denver II is not an IQ test; it is not a definitive predictor of future

adaptive or intellectual ability, nor is it designed to generate diagnostic labels (Frankenburg et al., 1992). Areas of developmental function assessed include person-social, fine motor-adaptive, language, and gross motor skills.

Standardization of the Denver II was conducted in 1988 by 17 screeners, trained to a high inter-rater reliability, who administered each item between 440 and 1,309 times. The sample population was a heterogeneous group of children from Colorado. The variables of sex, maternal education, ethnicity, and place of residence were analyzed to examine differences in age at which children passed an item. The standardization of the Denver II was the only method of determining validity used; justified by the test's authors' suggestion that the test could not be correlated with other tests because all tests are constructed slightly differently (Frankenburg et al., 1992). Glascoe, Byrne, Ashford, Johnson, Chang, and Strickland (1992), however, argue the following:

> Validation in comparison to widely accepted diagnostic tests is essential since this is the only method for determining how well each item performs in relation to diagnostic measures or other performance criteria (e.g., school success) and whether items actually measure meaningful aspects of child development (p. 1224).

The sensitivity and specificity of the Denver II is questioned by several researchers (Committee on Children with Disabilities, 2001; Glascoe, 2002; Glascoe et al., 1992). Despite the revisions that occurred to the original Denver Developmental Screening Test, the Denver II manual omitted information regarding the measure's validity. Researchers (Glascoe et al., 1992) argue that this omission is of concern

because there is little proof that the revision is a viable improvement from the original version of the test.

Glascoe (2002) reports that the Denver II's levels of both sensitivity and specificity are 44% dependent on how an administrator addresses questionable test scores. In an early study, researchers examined the validity of the Denver-II through comparison with the BSID-II, Kaufman Assessment Battery for Children, Stanford-Binet, 4th ed., and the Vineland Adaptive Behavior Scale (Glascoe et al., 1992). Similar to Glascoe's (2002) report of low level of specificity, the results of this study found a level of 43%. The level of sensitivity, however, was at a more acceptable level of 83%.

The low rates of specificity and the greatly varying rates of sensitivity provide NICU follow-up programs evidence that the Denver II may not be the most optimal instrument to screen children born prematurely. Although the Denver II is a widely utilized screening instrument, it does not offer items within each of the six developmental domains; the domains such as fine motor and adaptive are not assessed separately. Additionally, the administration of this instrument is lengthy and time consuming.

The most significant reason, however, that the Denver II is not appropriate with a NICU follow up Program population is that it lacks items within the first six months that assess children whose age at the time of the screening is still less than full term, particularly less than 28 weeks gestation. According to one prominent neonatologist, L. Blackmon (personal communication, January 26, 2003), there is a significant delay before Denver II items are manifest and the time range in which they appear is broad. The Denver II fails to use items that are appropriate for this age group's developmental status as well as for the quality of motor function, not merely the presence or absence of

landmarks. For example, within the gross motor domain, a one-month child will be screened to determine if there are equal movements and if he can lift his head; no information is obtained regarding the quality of these movements.

Bayley Scales of Infant Development, Second Edition (BSID-II)

Many studies have investigated the concurrent validity between instruments, such as the CAT/CLAMS, the Battelle, and the CDI, with the BSID-II (Kube, Wilson, Peterson, & Palmer, 2000; Provost, Crowe, & McClain, 2000; Wachtel, Shapiro, Palmer, Allen, & Capute, 1994). Revised from the original Bayley Scales in 1993, the BSID-II is considered as a widely accepted standard of infant development (Rossman, Hyman, Rorabaugh, Berlin, Allen & Modlin, 1994) and was used as the benchmark to measure concurrent validity in the current investigation. It is an in-depth evaluation that assesses the development of children from birth through 42 months of age by the administration of three subscales. The Mental Scale (MDI) examines memory, problem solving, conceptualization, language, and social skills. Fine and gross motor are assessed with the Motor Scale (PDI) and the Behavior Rating Scale examines arousal/attention, orientation/engagement, emotional regulation, and quality of movement. As the MDS does not examine behavioral state, the BSID-II's Behavioral Rating Scale will not be considered within the context of the present study.

Standardization of the BSID-II is based on a heterogeneous sample of 1,700 children, including 50 girls and 50 boys in each age group (Salvia & Ysseldyke, 1999, p. 619-621). The BSID-II classifies the developmental status of children as accelerated performance, within normal limits, mildly delayed, or significantly delayed. The classification status of a child is dependent on how much the MDI or PDI score (derived

from raw scores) deviates from the norm; one standard deviation (15 points) below the norm indicates a mild delay and two standard deviations below the norm indicates a significant delay. In addition to MDI and PDI scores, developmental ages can be derived from raw scores (Bayley, 1993). The developmental age is frequently used to determine whether the percentage of delay is great enough to qualify a child for early intervention services. Although classification status is the recommended method for the BSID-II to identify children with developmental delays (Bayley, 1993), developmental ages are often the criterion set forth by states (e.g., Maryland) for acquisition of early intervention services (Shackelford, 2004). Similar to other concurrent validation studies (Palisano, 1986; Provost et al., 2000) the developmental age equivalent was used as the criterion measure for this study.

Screening measures that have been examined concurrently with the BSID-II have been previously discussed. These instruments include the CDI (Diog et al., 1999), the Battelle (Ikle & Wittmer, 1995), the BINS Benish, 1998; Krug, 1998), and the Denver II (Glascoe et al., 1992). The Clinical Adaptive Test/Clinical Linguist and Auditory Milestone Scale (CAT/CLAMS) and Peabody Developmental Motor Scales (PDMS) are among the assessment instruments concurrently compared to the BSID-II. The CAT/CLAMS examines language, problem-solving abilities, and visual-motor skills for children ages birth to 36 months of age. Researchers (Kube et al., 2000) report a strong concurrent validity between the CAT/CLAMS and the BSID-II ($r = 0.89$, $p < .0001$). The results of Provost and colleges' (2000) concurrent validity investigation of the BSID-II Motor Scale and the PDMS with a sample of two-year-old children suggest that there is a very good to high correlation with age equivalent scores of the PDMS Fine Motor Scale

(r = .87) and with the PDMS Gross Motor Scale (r = .87). Further investigation of the correlation of these standard scores, however, demonstrated a poor to unacceptable correlation with both the PDMS' Fine Motor Scale (r = .64) and Gross Motor Scale (r = .49) (Provost et al., 2000).

Similarly to Provost and others (2000), Palisano (1986) examined the concurrent validity of the BSID-II's Motor Scale and the PDMS. Healthy infants born at full-term (n = 23) and prematurely (n =21) were assessed with both tests at 12, 15, and 18 months of age. Results of this investigation suggest a good to high correlation on age equivalent gross motor scores (r = .78 to r = .96) but yielded an unacceptable correlation with fine motor scores (r = .20 to r = .57). Palisano (1986) additionally found that with full-term infants the mean BSID-II quotients were significantly higher than the Peabody gross motor quotients. The findings of these studies suggest that although the BSID-II is the current benchmark for examining the concurrent validity of early childhood developmental assessment, one should be cognizant of the BSID-II's limitations and proceed with caution (Provost et al., 2000). Overall, however, the BSID-II provides a strong standard of assessment of infants and young children and comparison with this instrument will yield important information about tools, such as the MDS, which need to undergo thorough validation procedures.

Bayley's manual (1993) includes an examination of the BSID-II use with a population of infants born prematurely. The sample (N = 57) population's MDI score was approximately 4/5 of a standard deviation below (M = 88.6) the normative sample and the PDI score was 1 standard deviation below (M = 83.5) the normative sample.

These results suggest that despite a correction for prematurity, children born at less than 36 weeks gestation are more likely to perform below their same age peers.

Need for New Screening Measure

As illustrated in Table 3, not one of the twelve instruments discussed completely satisfied seven of the eight screening needs of children who are premature. The criterion of validity is not included in this table as all instruments included in the review had some validity measures. Meeting the following screening criteria is important to ensure that children in this population requiring further, in-depth assessment are identified as early as possible: (a) screens at least six developmental domains, (b) uses minimal materials, (c) easy and speedy administration, (d) tests infants whose adjusted age at time of screening is less than full term (40 weeks gestation), (e) provides hands-on assessment, (f) includes parent report, and (g) includes a professional evaluation of the quality of the child's performance.

While screening instruments including the ASQ, BABES, CDI, Healthy Steps and PEDS use minimal materials, are easy and speedy to administer, and provide opportunity for parent input, none provide hands-on assessment nor is it possible to assess children whose age at time of testing is considered to be less than full term. Of this group of instruments, only the CDI screen provides information on at least six developmental domains. These domains, however, include letter and number skills; not often applicable for screening very young children. The EIDP, Brigance Screen, BITS, and DISC screen at least six domains and are again easy and speedy to administer but do not consider parent report in determining results. Both the Brigance Screens and the DISC utilize more

than minimal materials and all three of these instruments again are unable to assess children less than full term.

Although the BINS combines parent report and direct administration, as it is adapted from the BSID-II, it uses more materials and is not as easy to administer as other instruments. Of the other instruments which combine direct administration and parent report, including the Battelle, Denver II comes the closest to effectively meeting the screening criteria. As with all other instruments, the Battelle and Denver II do not have the ability to assess children who are less than full term. Furthermore, as discussed, it is suggested that the Battelle does not provide enough items at each age interval (Ikle & Wittmer, 1995), and the Denver II's specificity and sensitivity has been challenged by many researchers (Committee on Children with Disabilities, 2001; Glascoe, 2002; Glascoe et al., 1992). Finally, none of these instruments included a professional evaluation of the quality of a child's performance.

Table 3

Criteria for Screening Infants Born Prematurely

Tool	Screens at least six* domains	Uses minimal materials	Easy and speedy to administer	Tests Adjusted Age < term	Hands-on	Parent Input	Eval. Of quality
Parent Report							
ASQ		X	X			X	
BABES		X	X			X	
CDI	X	X	X			X	
Healthy Steps		X	X			X	
PEDS		X	X			X	
Direct Administration							
BITS	X		X		X		
Brigance	X		X		X		
DISC	X				X		
EIPD	X	X	X		X		
Direct Administration & Parent Report							
Battelle	X	X	X		X	X	
BINS	X				X	X	
Denver II		X			X	X	
MDS	X	X	X	X	X	X	X

* Developmental domains: cognitive, fine motor, gross motor, language, social-emotional, adaptive

The Maryland Developmental Screen (MDS)

In response to the need for a screening instrument that meets the criteria described (see Figure 1) the MDS was created by Brenda Hussey-Gardner (2003) at the University of Maryland Medical System (UMMS). The instrument that the UMMS NICU Follow-Up Clinic used prior to the MDS' conception was the Maryland Premature Development Inventory (MPDI). Developed in 1987 by Renee Wachtel, the MPDI examined the developmental domains of gross motor, fine motor/adaptive, language/auditory, and visual/problems solving. The MPDI was designed to examine children whose age at the time of administration was considered to be less than full term. With this measure, it was possible to screen a child whose post conceptual age (PCA) was 36 – 40 weeks gestation, through the age of 36 months. The clinical observation and/or parent report of chronologically age appropriate items was recorded to determine a developmental quotient.

Several factors led the UMMS NICU Follow-Up Clinic to discontinue use of this instrument. Professionals who used this assessment strongly believed that it did not adequately meet the screening needs of a NICU follow-up program's population (personal communication with B. Hussey-Gardner, April 28, 2005). Items used to determine skills were provided monthly for the first year; after one year of age, items were administered every two months, and after two years of age items were only provided at six-month intervals. There was concern that this widening gap in test items delayed some children from being identified as requiring further in-depth assessment. In addition, the MPDI did not screen the six developmental domains separately. Finally, the MPDI was not subjected to rigorous validity procedures. Consequently, this screening

instrument was no longer used at this clinic and the new MDS replaced the instrument's function.

The purpose of the MDS is to "identify infants and toddlers whose developmental progression may be delayed and would benefit from more in-depth evaluation" (Hussey-Gardner, 2003). As with other screening instruments the MDS does not replace formal evaluation/assessment procedures nor is it intended to elicit diagnosis. Unique to this new measure, it is designed to be used to screen the development of children 30-32 weeks gestation through 39 months of age. The domains screened with the MDS include cognitive, social-emotional, adaptive, gross motor, speech and language, and fine motor development.

Similar to the BSID-II, the MDS may be characterized as "theoretically eclectic" (Bayley, 1993, p. 2). The 181 test items were compiled from screening instruments, assessments, curriculums, and a broad cross-section of infant and child research. As with the BSID-II, this approach has yielded a representation of diverse viewpoints with no particular theory driving the content. Through a validation process in which professionals reviewed domains for which they had expertise, MDS test items were placed at the correct age and that the items represented key milestones for that age. Nine professionals have reviewed test items within their developmental expertise. Professionals who reviewed the MDS included a developmental pediatrician, neonatologist, physical therapist, occupational therapist, speech and language therapist, psychologist, and special educator. Each individual examined the domain in which they were most knowledgeable and provided a critique on the appropriateness of selected assessment items. The professionals verified that each item was placed at the correct age and that the items were

representative of key milestones. Modifications of the MDS were made based on the reviewers recommendations until consensus was achieved (Hussey-Gardner, 2003). After one year of piloting, additional changes were made to the MDS. For example, the six month cognitive skill of 'lifts cup by handle" was omitted because most children were observed to reach around the handle to grasp the cup. Many parents suggested that sippee cups do not have handles and therefore the children were not aware of their function.

The MDS is scored by recording whether each item is observed as pass by the examiner (O), reported as a pass by the parent (R), or neither observed nor reported as a pass (N). Unique to this measure, the quality of every item is additionally evaluated by the professional and recorded as either a typical (T) or atypical (A) performance. Based on the O, R, and N scores in each age range, a child is considered to pass the screen, be suspect, or fail that particular age range. A child passes the screen if all items are scored either O or R. A child is considered suspect if one item is scored N but all other items are scored O or R. A child fails the screen if a N is scored in the current and preceding age range in one or more domains. The examiner also indicates whether the quality of the child's performance is typical and also notes whether the parent feels that the child performed as they expected. If a parent reports that the child did better than expected a '+' is noted next to the "N"; a '—' indicates that the child's performance was not as optimal as expected. An example is provided in Figure 2.

Figure 2.

Example of MDS Items

Cognitive		Language		Social-Emotional		Adaptive		Gross Motor		Fine Motor		Results
Tracks ring 30^{o}	O R A N	Throaty noises	O R A N	Quiets to voice/face	O R A N	Can maintain alertness 1 hour	O R A N	Head up in prone	O R A N	Hands fisted	O R A N	P S F T A Y N

Key: O = Observed, R = Reported, A = Atypical quality, N = Neither observed nor reported. P = Pass: Child receives O or R for all items in age, re-screen at next visit, S = Suspect: Child receives N in one or more domains but receives O or R for all items in that domain at preceding age range, re-screen at next visit, if two consecutive suspects refer to developmental team, F = Fail: Child receives N in current & preceding age range in one or more domains, refer to developmental team. T = Typical quality of performance & no A's marked in row, A = Atypical quality of performance with at least one item marked A in row, refer as needed. Y = Yes child's performance was as parent expected, N = No child's performance was not as parent expected (+ indicates better than expected, - indicates not as good as expected)

In the setting in which it is currently used, the NICU Follow-Up clinic, the professionals (e.g., neonatologists) who administer the MDS have training and qualifications which should enable them to demonstrate reliability in identification of atypical development. In another setting, the reliability of the identification of an atypical quality may be problematic because the administrators may not have the expertise to identify potential problems. Clinical judgement will not be sufficient in other environments and therefore the MDS manual is currently under revision so that quality of performance on test items will be clearly defined.

As the survival rate for infants born prematurely increases, researchers suggest that risk for these children developing disabilities increases as well (Berger, Holt-Turner, Cuppoli, Mass, & Hagerman, 1998; Linden et al., 2000; Resnick, Eyler, Nelson, Itzman

& Bucciarelli, 1997). Therefore it is important to frequently screen and assess this population's developmental status. Few studies, however, were found that examined any screening instrument's validity when used with a population of children born prematurely. In addition to identifying children in need of early intervention as early as possible, accurate developmental screening reminds professionals to observe and record clinical observations and assists in appropriately allocating health care services (Committee on Children with Disabilities, 2000; Glascoe et al., 1992; McLean et al., 1996; Sonnader, 2000).

The administration of a developmental screening test to infants and toddlers born prematurely occurs through parent report, direct report, and a combination of parent report and direct administration. Twelve screening instruments were reviewed to determine which, if any, were appropriate to measure the developmental status of infants born prematurely. This review yielded no validated developmental screening instruments which met the criteria of (a) six developmental domains, (b) use of minimal materials, (c) ease and speed in administration, (d) ability to test young infants with adjusted age of less than term, (e) hands-on assessment, (f) parent involvement in the screening process, (g) professional evaluation of the quality of the child's performance and (h) valid screening measure of developmental status.

In response to the need for a new screening instrument, the MDS was created. As with other screening instruments, the MDS is neither intended to replace formal assessment procedures nor to elicit diagnosis. Rather, the purpose of this instrument is to identify children whose development may be delayed and who would benefit from more in-depth developmental evaluation. The MDS screens children, whose ages range from

30-32 weeks PCA through 39 months, for potential delays in cognitive, language, adaptive, social-emotional, gross motor, and fine motor development. Although the MDS appeared to meet an unmet need for accurately screening infants and toddlers born prematurely, it had not yet to be subjected to rigorous validation procedures. To respond to this need this study was conducted to determine the concurrent validity of the MDS when compared to the "gold standard" (Sattler, 1992, p. 321) of developmental assessments, the Bayley Scales of Infant Development, Second Edition (BSID-II), with infants and toddlers who were born prematurely. To address this issue the following three research questions were asked:

1. What is the sensitivity of the MDS when concurrently administered with the BSID-II?
2. What is the specificity of the MDS when concurrently administered with the BSID-II?
3. What is the overall hit rate and chance hit rate of the MDS when compared to the BSID-II?

CHAPTER III

Research Methodology

The purpose of this study was to investigate the concurrent validity of the Maryland Developmental Screen (MDS) when compared to the Bailey Scales of Infant Development, Second Edition (BSID-II) in a population of infants and toddlers born prematurely. To address this issue the following three research questions were asked:

1. What is the sensitivity of the MDS when concurrently administered with the BSID-II?
2. What is the specificity of the MDS when concurrently administered with the BSID-II?
3. What is the overall hit rate and chance hit rate of the MDS when compared to the BSID-II?

A description of the site, participants, research methodology, instrumentation, field procedures, data collection and recording, data processing and analysis, and methodological assumptions is provided.

Site

This study was conducted at the University of Maryland Medical Systems (UMMS) NICU Follow-Up Clinic. The purpose of the clinic is to: (a) coordinate neurodevelopmental screening from birth to age three, (b) provide multidisciplinary support, including speech and language therapy, social work, physical therapy, occupational therapy, psychology, and early intervention service coordination for children who have been discharged from the NICU, and (c) provide consultation on unresolved medical problems following NICU discharge. The clinic's two components,

screening and assessment, coordinate with primary care to provide optimal developmental and medical monitoring. Children who are born at 36 weeks gestation or less receive an age adjustment for their level of prematurity until their adjusted age is 12 months. At either their chronological age or adjusted age, the MDS and the medically based neurodevelopmental examinations are administered. The MDS screens the domains of cognition, fine motor, gross motor, adaptive, social-emotional, and language development. The neurodevelopmental examination assesses a child's "neuromuscular status (e.g., tone), primitive reflexes (e.g., moro), developmental reflexes (e.g., grasp), protective reactions (e.g., downward parachute), involuntary movement (e.g., tongue thrust), deep tendon reflexes (e.g., ankle), mobility (e.g., sit to crawl), fine motor (e.g., transfer), and oral reflexes" (Hussey-Gardner, 1995, p. 56). Although the medical neurodevelopmental exam is a component of each clinic visit, the results were not collected, nor part of this study's analysis. The purpose of this study was to investigate the validity of the MDS with infants and toddlers born prematurely. It was determined that by including this information, the likelihood of generalizability, thus validity, of this study would decrease because in different settings (e.g., early childhood centers) neonatologists are often not present to collect medical information.

A child who passes both the MDS and neurodevelopmental screens attends routine appointments at the NICU Follow Up Clinic at 3, 6, 9, 12, 18, 24, 30, and 36 months of age and is screened with both exams at each appointment. Although not considered a routine appointment, a child may also been seen at 39 months of age if they have missed their 36 month appointment or if the clinic staff make new service

recommendations or provide a new diagnosis at 36 months of age. Children who are 24 months of age also receive the BSID-II as part of routine care.

If a child's development is suspect or delayed, screening or evaluation may occur at more frequent intervals. In addition, if the results of the MDS indicate that a child's development is suspect or delayed, one or more developmental staff member(s), whose expertise matches the child's needs, will conduct a more in-depth assessment. Based on a child's individual needs, the developmental team staff members who may conduct assessments include an occupational therapist, physical therapist, speech and language therapist, psychologist, and special educator. If the child fails the MDS or is suspect on two consecutive administrations, the child is referred to be seen by the developmental pediatrician and one or more members of the developmental team. Once the developmental pediatrician evaluates a child, he or she is no longer screened through the MDS, unless all developmental concerns are resolved. For the purpose of this investigation, however, all children whose families agree to participate, regardless of whether they routinely received the MDS or were followed by the developmental pediatrician, received both the MDS and BSID-II.

Participants

The participants in this study were infants and young children who were followed by the UMMS NICU Follow-Up Clinic. A list of factors that serve as eligibility criteria for attending the NICU Follow-Up Clinic and that place children at-risk for developmental delay is provided in Table 4. All infants who attend the NICU Follow-Up Clinic were inpatients in the UMMS NICU or one of six other Maryland NICUs (Hussey-Gardner et al., 2002) at birth and had been discharged based on their medical stability.

The children who attend the UMMS NICU Follow-Up Clinic are representative of the total population of premature infants born in Maryland, in terms of SES, race/ethnicity, and diagnoses. There are approximately 650 infants who are admitted to UMMS NICU per year. These infants represent families from all jurisdictions in Maryland, as well as from surrounding states.

Table 4.

UMMS NICU Follow-Up Clinic Criteria Conditions

- Birth weight < 1,500 grams
- Gestational age <32 weeks
- Apgars < 5 at 5 minutes
- Intrauterine growth retardation
- Intraventricular hemorrhage >Grade II
- Congenital infection
- Congenital anomalies
- Ventriculomegaly
- Porencephaly
- Hydrocephalus
- Microcephaly
- Hypoxi-ischemic encephalopathy
- Periventricular leukomalacia
- Seizures
- Meningitis
- Broncopulmonary dysplasia
- Hearing impairment
- Vision impairment
- Abnormal neurolgic exam at discharge

Once discharged from the UMMS NICU, most children attend the UMMS NICU Follow-Up Clinic. Additionally, the clinic receives referrals from six other NICUs located in Maryland. In 2004, 648 children attended the UMMS NICU Follow-Up Clinic for one or more visits. The ethnicity of these children was as follows: 361 (55.7%) African American, 221 (34.1%) Caucasian, 13 (2%) Hispanic, 7 (1%) Asian, 3 (.5%) Native American, and 43 (7%) Other/Unknown. Although the majority of families who

attend the NICU Follow-Up Clinic reside in Baltimore City, families travel to the clinic from all jurisdictions in Maryland. The geographic representation of the clinic population in 2004 was as follows: 517 (78%) Greater Baltimore, 49 (8%) Capital Region, 13 (2%) Eastern Shore, 31 (5%) Southern Maryland, and 19 (3%) Western Maryland. The families varied in respect to socio-economic status, ethnic group, and religion. There is also diversity in family structure, including two-parent and single-parent families, grandparent caregivers, and foster/adoptive parents.

Approximately one-third (105) of all children who attend the clinic receive early intervention services from one of Maryland's 24 counties' Infants and Toddlers Programs. In the state of Maryland, if a child demonstrates a developmental delay of 25% or more in one or more areas of development, is developing atypically, or has a medical condition in which there is a high probability for developmental delay, early intervention services are provided to the family at no cost (Maryland Infants and Toddlers Program, 2000). Services may include, but are not limited to, special instruction, speech and language therapy, occupational therapy, physical therapy, and early intervention service coordination.

Caregivers were recruited for the study as they were waiting for their infants' appointments in one of the NICU Follow-Up Clinic examination rooms. A sample of 81 children who were born at less than or equal to 36 weeks gestation and whose chronological ages ranged between 3 and 38 months were recruited for participation in this study. To have a sample representative of all age groups, the recruitment for the sample was stratified to ensure that there were at least 10 children represented for each visit level (i.e., 3, 6, 9, 12, 18, 24, 30, and 36 months of age). Exclusion criteria for the

proposed study included children whose caregivers did not speak English. Any child born at 37 weeks gestation or greater was also excluded, as these children were not considered to be premature. Those children whose adjusted age was less than or equal to 16 days old were additionally excluded because the BSID-II is not capable of assessing children this young. Finally, children who visited the clinic at 38 months of age or older were excluded as these visits were not considered to be routine.

Instrument Selection

The present study examined the validity of the MDS by determining sensitivity and specificity identification rates of children born prematurely who were and were not developmentally delayed as measured by BSID-II assessments. Although rarely stated in academic literature, clinically, the BSID-II is considered to be a gold standard in early childhood assessment. Sattler (1992) describes this instrument as "by far the best measure of infant development and provides valuable information about patterns of early mental development" (p. 321). The BSID-II is the most commonly used standard against which most other screening instruments have been compared. The BSID-II has been used to examine the concurrent validity of the Harris Infant Neuromotor Test (Harris & Daniels, 2001) and the Bayley Infant Neurodevelopmental Screener (Leonard et al., 2001) with populations of infants born prematurely. Researchers have also used the BSID-II to establish validity of screening instruments when used with infants born at full term. These instruments include the Denver II (Glascoe et al., 1992), the Child Development Inventory (Diog et al., 1999; Shoemaker et al., 1993) and the Brigance Screens (Glascoe, 2001). Finally, the BSID-II has been used as the standard to which other in-depth assessments including the Clinical Adaptive Test/Clinical Linguist and

Auditory Milestone Scale (Kube et al., 2000; Watchel et al., 2000) and the Battelle (Ikle & Wittmer, 1995) are compared. Unlike the MDS, the BSID-II does not specifically address the six developmental domains; rather, the manual reports that domains are assessed directly by various items in the Mental Developmental Index (MDI) and Psychomotor Developmental Index (PDI). This issue was explored through the analysis.

Due to wide use of the Denver II, initial consideration was given to comparing the MDS to the Denver II. However, a decision was made to not additionally compare the MDS to sensitivity and specificity rates of the Denver II. Researchers have found low rates of both sensitivity (44% to 83%) and specificity (43% to 44 %) when relating the Denver II to other assessments including the BSID-II (Glascoe, 2002; Glascoe et al., 1992). As indicated earlier, a 70% sensitivity level and a 70-80% specificity level are acceptable rates for developmental screening (Sonnader, 2000). Although the Denver-II's sensitivity level was 83% in one study (Glascoe, 1992), the range (44% to 83%) among studies is too great. The low sensitivity and specificity rates demonstrated, as well as the inappropriateness of the Denver II's test items for screening children less than six months of age, resulted in the omission of the use of this screening tool in the present study.

Methods and Procedures

Prior to initiating data collection, Institutional Review Board (IRB) approval was obtained from the University of Maryland Medical School (UMMS) and the University of Maryland, College Park. The student researcher was trained to administer both the MDS and the BSID-II. Reliability was established among administrators prior to initiation of data collection. Reliability of the MDS was established by comparing the neonatologists' and student researcher's administration and scoring of 12 (15% of the

sample) administrations with the clinic coordinator, who was responsible for training the UMMS NICU Follow Up Clinic staff to use the MDS. Reliability of BSID-II index scores was established between the staff psychologist, who routinely administered the BSID-II, and the student researcher by scoring 12 MDI (15% of the sample) and 12 PDI assessments.

The protocol for establishing reliability on both the BSID-II and the MDS was the same. Children were sequentially selected at random so that there were four children represented for each chronological age range of (1) 3 through 12 months, (2) 13 through 25 months, and (3) 26 through 37.5 months. The student researcher and psychologist alternated administration and scoring of the BSID-II. Neonatologists and the clinic coordinator simultaneously scored the MDS while alternating administration of the screen. The correlation of the total score of each test administered resulted in a reliability coefficient for the scorers obtained by the Cohen's Kappa. An interrater reliability exceeding .80 for the MDS and .90 for the BSID-II respectively was desirable as determined by the agreement between scores from tests administered by the examiners prior to data collection. Additional interrater reliability was re-established through monthly periodic checks.

Upon agreement to participate, the families were asked to sign a consent form (see Appendix B). During clinic visits, a neonatologist administered the MDS to children whose development is typical or suspect at 3, 6, 9, 12, 18, 24, 30, and 36 months of age. If a child did not receive the MDS as part of routine care (i.e., received follow-up by the developmental pediatrician) the student researcher administered the MDS. Data was not collected on children whose corrected age was less than 16 days old because the BSID-II

did not have the ability the assess children this young. All children who are 24 months of age, as well as those whose cognitive development was found suspect or delayed by the MDS, were routinely administered the BSID-II by the clinic's psychologist. When a child was not scheduled to receive the BSID-II as part of routine care, the student researcher administered the assessment. To preserve the independence of two administrators, no one individual administered both the MDS and BSID-II to any one child.

A random table of numbers was to be used to determine whether the MDS or the BSID-II was administered first. This, however, did not occur, and is an acknowledged limitation of this study. The nature of the clinical environment, time constraints, or parent preference made it difficult to control the order of assessments. Instead, the order of the testing was most often dictated by the neonatologists' availability. For example, if a neonatologist was examining another NICU Follow-Up Clinic patient, then the BSID-II would be administered first; if the neonatologist was ready to examine the child in the study, the MDS would be administered first. All testing procedures occurred in one of the UMMS NICU Follow-Up Clinic's available examination rooms. Standardization of the testing environment was accomplished by ensuring that each room included a child's table and chairs, and a floor mat. The data collected was kept in a locked file box until the appropriate sample population was gathered. All analysis was completed through the use of the Statistical Package for the Social Sciences.

Administration of the MDS

The clinic coordinator established reliability with neonatologists prior to the initiation of data collection. Reliability was established with 1.0 agreement obtained using a Cohen's Kappa calculation with 12 subjects whose data were not included in the study's sample. The majority of children were tested with the MDS by a neonatologist as part of routine care. Children who did not receive the MDS as part of routine care were given the screen by the clinic coordinator. These children included those who had already been identified as developmentally delayed, and therefore were most often examined by the developmental pediatrician and received a more in-depth assessment such as the CAT/CLAMS or the BSID-II.

As there is dispute in the literature as to when age adjustment should be discontinued, age correction followed the MDS manual's protocol. (Brenbaum & Batshaw, 1997; Rickards, Kitchen, Doyle, & Kelly, 1989; Wilson & Michaeleen, 2004). The administrator adjusted the child's age in order to give the screen using the most appropriate test items. The MDS was administered at the child's adjusted age if the child was born at or less than 35 weeks gestation and the age continued to be adjusted until the child's adjusted age was 12 months old. The adjusted age was computed by subtracting the number of weeks of prematurity from the child's chronological age. Rationale for adjusting a child's age until the adjusted age of 12 months was provided in the MDS manual (Hussey-Gardner, 2003):

> (1) The earlier a child is born, the longer the child needs to 'catch-up', and (2) adjustment beyond the age of 12 months may result in a child not being referred for further evaluation and early intervention services at an early enough age (p. 2).

Once an adjustment for prematurity was calculated, all items in an appropriate age row were administered. A score of 'O' (observed), 'R' (parent report), or 'N' (not observed or reported) was given for each item. Additionally, each item was scored either 'A' (atypical) or 'T' (typical) to reflect the professional's evaluation of the quality of the child's performance. If an 'N' was scored within any domain, the examiner tested items at an earlier age. For example, if a child was screened at 21 months of age and neither observed by the administrator nor reported by the parent as capable of the "unzips zipper" skill, then the 18-month skill of "uses a spoon with little spilling" was also administered (see Figure 3). A child passed the screen if an 'O' or 'R' was scored for all items within the appropriate age range. A child who passed continued to be re-screened at the NICU Follow Up Clinic's routine scheduled appointments.

A score of 'N' in one or more domains at the appropriate age range with an 'O' or 'R' for items in the preceding age range yielded a suspect result. The child's development was considered to be questionable and anticipatory guidance was provided to the parent. The child was re-screened within three months and if she or he scored fail or suspect on the consecutive screen, a referral for a developmental assessment was made. A child failed the MDS if he or she received an 'N' in the appropriate and proceeding age range in one or more domains. A failure on the MDS indicated that a child may not be developing skills typical for children of the same age and further assessment was necessary.

A more in-depth evaluation in the domain of concern, should be conducted as soon as possible. For example, if a child failed the language domain, a speech and language pathologist should be a participant of the subsequent evaluation and based on

assessment results, appropriate early intervention referrals should be completed. The MDS additionally consists of two qualitative components. First, the professional administrator indicated whether or not the child's development presented with typical or atypical quality. An example is that a child is walking at an appropriate age but has an atypical quality because he or she drags the left leg. If any individual test item was scored atypical, then the overall test was scored atypical. Second, the parent had the opportunity to report whether or not the child performed as expected. If the child did better than expected, anticipatory guidance and parent education was tailored to the parent's understanding of development. If the child did not perform as expected it may have indicated that there may be a better time to evaluate the child (e.g., not during not nap time); however, the screening results were still considered valid as parent report is used to score the screen as well as professional observations.

Figure 3

Example of Scoring the MDS

Screen Age	Adaptive	O R A N	Results
12	Takes off hat, shoes socks	O R A N	P S F – T A Y N
15	Uses spoon with little spilling	O R A N	P S F – T A Y N
18	Unzips zipper	O R A N	P S F – T A Y N
21	Wipes nose if given a tissue	O R A N	P S F – T A Y N

Administration of the BSID-II

Prior to initiating this study, reliability on the BSID-II was established. The researcher and psychologist reached an agreement of 1.0 using Cohen's Kappa with 12 subjects (15% of the sample). The agreement between evaluators was based upon whether a subject's developmental age was less than 25% delayed or 25% delayed or greater. Data obtained for these subjects were not included in the sample population.

Once reliability was established between administrators, each child whose family was willing to participate in the study was administered the mental developmental index (MDI) and the psychomotor developmental index (PDI) of the BSID-II by either the psychologist or the student researcher. Either the psychologist or the student researcher administered the assessment in a patient examination room at the clinic. The age range in which the BSID-II was administered was determined by a calculation of the child's age with an adjustment for prematurity. To administer the BSID-II the examiner administered items that were scored as credit (C) or no credit (NC). Scores of the MDI and PDI are standardized so that there is a mean of 100 and a standard deviation of 15. Raw scores were obtained by adding the number of items for which the child received credit to all the items below the basal item. Raw scores were determined based on the standardization tables provided for each age in the test manual. The raw scores were then used to determine an age equivalent score to indicate whether the child was 25% delayed.

Demographic information

Demographic information was gathered regarding the child including date of birth, gestational age, sex, race, birth weight, diagnosis (while in the NICU and current), family member(s) present at the exam, and whether the child was involved in the early intervention system (See Appendix C for the Demographic Questionnaire). Family information was also gathered, including Maryland county of residence as well as and the parent's age and marital, employment, and income status. The first nine questions on the demographic form were obtained through a chart review and the following eight were obtained through parent interview. Descriptive statistical procedures conducted on the demographic data included measures of central tendency (mean) and measures of variance (range and standard deviation). This information was used to analyze variables that may have contributed to trends in the rate of referrals of children who reportedly require more in-depth assessment.

Data Analysis

An investigation of the concurrent validity of the MDS was conducted by analyzing the data to determine the instrument's rates of sensitivity and specificity as determined by the criterion, the BSID-II assessment. This was accomplished by examining the true positive, false positive, true negative, and false negative rates of MDS scores as measured by BSID-II assessment results. Demographic data was also analyzed to identify potential referral trends and to provide descriptive information regarding the sample.

The Hit Rate Model

The Hit Rate Model (Lichtenstein and Ireton, 1984), which is descriptive in nature, was used to determine whether MDS scores concurrently predicted BSID-II scores, thus identifying which children were in need of early intervention services. The hit-rate model "provides a methodological tool for determining the validity of a screening process or for comparing one prospective screening process with another" (Lichtenstein & Ireton, 1984, p. 230). This model summarizes the relationship between the outcome of a screening measure and the actual status of an individual in a given population (Litchtenstein & Ireton, 1984). In the case of this study, the outcome of the screening measure was a child's classification on the MDS and the actual status referred to the classification outcome (i.e., pass or fail) on the criterion measure, which was the BSID-II.

Despite the fact that the BSID-II has two classification categories (< 25% delayed or ≥25% delayed) and the MDS has three classification categories (pass, suspect, or fail) the two instruments were compared. The three MDS categories of pass, suspect, and fail were collapsed into binary categories for analysis. Children who received a suspect classification when screened with the MDS were grouped with children who passed the screen. Justification for grouping pass/suspect together was that children were not expected to exhibit developmental delays if they receive a screening result of suspect. In a clinical setting, these children would be monitored more closely but not referred for more in-depth evaluation.

Results of the BSID-II were collapsed into binary categories to indicate a developmental delay versus a non-delay using a 25% delay criteria. The decision to group children into the categories of 25% or greater delay and less than 25% delay was made

because in the state of Maryland, children are eligible for early intervention services if they demonstrate a delay of 25 % or greater. Despite the considerable variability among states in the quantitative description of developmental delay and the wide range of the level of delay that is required by states for eligibility for early intervention services, a 25% delay is one of the most common quantitative criteria used (Shackelford, 2004). This fact provided evidence that the results of this study would be applicable to many states that follow the same eligibility guidelines as Maryland. Age equivalents associated with raw scores provided by the BSID-II manual were used to determine a 25% delay or non-delay. For example, a 24-month-old child whose age equivalent was 17 months on the mental or motor scale would be considered to be developmentally delayed on the BSID-II.

Children who were screened by the MDS were categorized by two outcomes. They were either categorized as screening positive, meaning that they were at a high risk for developmental delay and needed to be referred for further assessment through the BSID-II or as screening negative, therefore considered low risk and did not need to be referred. Following Lichtenstein and Ireton's guidelines (1984), based on these two screening outcomes there were four possible results for each child tested:

(1) True Positive: A child was referred by the screening procedure and failed the BSID-II (the MDS was accurate).

(2) True Negative: A child was not referred by the screening procedure and did fail the BSID-II (the MDS was accurate).

(3) False Positive: A child was referred by the screening procedure but did not fail the BSID-II (the screen yielded an over-referral).

(4) False Negative: A child was not referred by the screening procedure but failed the BSID-II (the MDS inaccurately yielded an under-referral).

When taking into account the four possible screening results, the indices of sensitivity, specificity, overall hit rate, and chance hit rate were determined (Speece & Cooper, 2004). Sensitivity, in terms of hit rate analysis, refers to the percentage of subjects correctly detected at screening as having a developmental delay. Specificity is the percentage of subjects correctly passed by the screening due to the absence of a developmental delay (Speece & Cooper, 2004).

The rate of false positives was calculated to provide evidence of the level of specificity of the MDS. For example, a high rate of false positives would yield lower specificity and lower overall hit rates. The rate of false negatives was related to the sensitivity of the MDS. A high rate of false negatives would yield lower sensitivity and lower overall hit rates (Speece & Cooper, 2004). The overall hit rate, defined as the "proportion of accurate screening decisions out of the total number of screening decisions" (Lichtenstien and Ireton, 1984) was calculated. The chance hit rate was additionally calculated to address the potential for a deceptively high overall hit rate. (Speece & Cooper, 2004). The chance hit rate indicates how many subjects would, by chance alone, obtain the same screening outcome. Chance hit rates were considered when analyzing this data because although the majority of children (67%) will develop age appropriately, 33% children born at VLBW are at an increased risk for neurological problems (Nadeau et al., 2001). Rates of developmental problems with preschoolers fall

in the range of 5-10% of the population and the rate for a high-risk population, such as infants and toddlers born prematurely, is 1.33 to 2.5 times higher (Litchenstein & Ireton, 1984).

Analysis of MDS and BSID-II Data

To be considered valid, the MDS, like other screening procedures, must have acceptable levels of sensitivity and specificity. The sensitivity of the MDS refers to the instrument's ability to identify correctly those children who are determined, through more in-depth assessment, to have significant (>25%) developmental delays, and who will ultimately require early intervention services. The specificity of the MDS refers to the test's ability to identify those children who are neither significantly delayed (<25%) nor in need of early intervention services. As with any other screening instrument's sensitivity and specificity rates, an increased level in either rate will result in a risk of a decreased level in the other. In other words, an instrument that is highly sensitive may over-identify children; all appropriate referrals for in-depth assessment may be made but children who are not delayed may be incorrectly identified, thus subjecting them to unnecessary assessment. Conversely, if an instrument is highly specific, while children not in need of in-depth assessment will not be subjected to testing, the risk is great that children who are developmentally delayed may not be identified.

The decision whether to use an instrument which is more sensitive or specific is dependent on several factors. According to Lichtenstein and Ireton (1984), a common guideline is that it is more grave for an instrument to be less sensitive as this may lead to under-referral to needed early intervention services. Local priorities and policies, however, often drive the decisions whether to use a more sensitive or specific instrument.

Furthermore, the emotional, clinical, and financial costs in coping with classification errors of under- and over-referral must be considered.

It was desirable for the MDS to err on the side of increased sensitivity rather than specificity so that all children who are potentially eligible for early intervention services receive them in an appropriate manner. Families routinely attend the NICU Follow-Up Clinic in order for their child to receive medical and developmental follow-up related to their prematurity. Clinicians attend the clinic and receive payment for their time regardless of the number of children seen. Although children who receive false positive results on the MDS will be subjected to timely in-depth assessments, often the assessments occur on the same day as the screening or at the next routine appointment, therefore there minimizing time constraints for the parents. Furthermore, in an effort to make the clinic experience optimally family-focused, the clinic's coordinator informs families of the potential of any screening instrument to over-identify children at risk.

To determine the sensitivity and specificity rates of the MDS the hit-rate model was utilized. Results of the MDS and the BSID-II were examined to determine the rate at which the MDS correctly referred children to receive the BSID-II, the in-depth assessment. As illustrated in Table 5, each subject's score on the MDS was recorded as falling into one of the four following screening categories:

A. Accurate Referral/True Positive - results of MDS and BSID-II both indicated developmental delay
B. Under-Referral/False Negative – results of MDS indicated no delay; BSID-II indicated delay

C. Over-Referral/False Positive – results of MDS indicated delay; BSID-II indicated no delay

D. Accurate Non-referral/True Negative – results of both the MDS and BSID-II indicated no delay

Table 5

MDS Screening Outcome Categories

MDS	BSID-II	
	Developmental Delay (>25%) (+)	No Developmental Delay (≤25%) (-)
Refer Fail (+)	A Accurate Referral/ True Positive	C Over-Referral/ False Positive
Do not refer Pass/Suspect (-)	B Under-Referral/ False Negative	D Accurate Non-referral/ True Negative

Adapted from Lichtenstien & Ireton (1984)

The rate of sensitivity (A/A+B) of the MDS was calculated by dividing the number of accurate referrals (true positives) by the number of accurate referrals plus the rate of under-referrals (false negatives). The specificity rate (D/C+D) was calculated by dividing the number of accurate non-referrals (true negatives) by the rate of over-referrals (false positives) plus the number of accurate non-referrals. The overall hit rate (A+D/N) was determined by dividing the number of accurate referrals (true positives and true negatives) by the total sample. The chance hit rate ($p^2 + q^2$) was established by examining the prevelance (p) of the condition, defined as all children who were identified

as delayed by the BSID-II (true positives and false positives) and the rate of the condition in the population (q = 1-p). Establishing the hit rate patterns of the MDS (i.e., whether there is higher sensitivity or specificity, etc.) will provide evidence to help programs decide if use of this screening tool is compatible with their objectives and goals.

Chapter IV

Results

The study was designed to determine the validity of the Maryland Developmental Screen (MDS) when concurrently administered with the Bayley Scales of Infant Development, Second Edition (BSID-II) as measured through indices including sensitivity, specificity, overall hit rate and chance hit rate. Specifically the research questions were:

1. What is the sensitivity of the MDS when concurrently administered with the BSID-II?
2. What is the specificity of the MDS when concurrently administered with the BSID-II?
3. What is the overall hit rate and chance hit rate of the MDS when compared to the BSID-II?

The two measures were independently administered to a sample of infants and toddlers born prematurely. In addition, relevant demographic information regarding the subject and family was obtained. The study was conducted on Wednesdays at the University of Maryland Medical System's (UMMS) NICU Follow-Up Clinic for approximately nine months (2/18/04-11/3/04).

A total of 103 families were approached to participate in this study. Thirteen families did not wish to participate and provided the following reasons: time constraint (n = 5), work schedule (n = 1), commute/traffic (n = 1), expiring parking meter (n = 1), child too tired or upset (n = 4), or caregiver's health (n = 1). Of the 90 families agreeing to participate, testing was not completed on seven children as two fell asleep and five

were unable to complete testing due to lack of cooperation (e.g., crying). Two additional subjects were excluded from the study as one child was born at full term (and not preterm as required by inclusion criteria) and the other was not tested at the correct age. The final sample for this validation study included 81 children and their families. The sample was stratified to represent the NICU Follow-Up Clinic's visit schedule: 3, 6, 9, 12, 18, 24, 30, and 36 months of age. Caregivers of subjects were sequentially approached and recruited as to meet the needs of the stratification (See Table 6).

Table 6

Stratification of Sample by NICU Follow-Up Clinic Visit Month

Visit	N	Percent
3 month	10	12.3%
6 month	9	11.1%
9 month	13	16.0%
12 month	10	12.3%
18 month	10	12.3%
24 month	9	11.1%
30 month	11	13.6%
36 month	9	11.1%
Total	81	100.0%

Demographic Information

Demographic information was collected with the Demographic Questionnaire (see Appendix C) via chart review and orally administered questions. Following completion of consent procedures, parents were asked to respond to the questionnaire while in the examination room either prior to or following test administration. The researcher also reviewed the child's chart to gain NICU diagnoses and other relevant early intervention

information. The following infant characteristics and family characteristics were examined.

Infant characteristics. The sample included (55.6%) male and (44.4%) female infants and toddlers. The subjects were predominately African American (66.7%), with the following ethnicities races reported: Caucasian (28.4%), Hispanic (1.2%), Bi-Racial (2.5%), and Other (1.2%). The gestational ages of the sample ranged from 23 weeks to 35 weeks gestation with a mean of 29 weeks gestation and a standard deviation of 2.63. Ten percent of the sample were twins. The birth weights of the sample ranged from 542 grams to 2,634 grams, with a mean birth weight of 1,198 grams and a standard deviation of 390.36. Additional information regarding the breakdown of the sample's gestational ages and birth weights is provided in Table 7. The children's ages ranged from 4 months of age through 37 months of age. When adjustments for prematurity were made according to the NICU Follow-Up Clinic protocol, adjusted ages ranged from 18 days through 37 months of age.

Table 7

Infant Characteristics (N = 81)

Gestational Age	N	Percent of Sample	Birth weight	N	Percent of Sample
23-26 weeks		21%	Extremely Low Birth Weight (<1000 g.)	24	30%
27-30 weeks		41%			
31-35 weeks		28%	Very Low Birth weight (>1001g <1500 g.)	40	49%
			Low birth weight (> 1501g. <2500 g.)	17	21%

The sample's NICU history was examined to determine frequency of other conditions related to high probability of developmental delay (see Table 8). As illustrated in Table 5, NICU histories included the following diagnoses: birth weight less than 1,200 grams, bronchopulmonary dysplasia (BPD), patent ductus arterious (PDA), necroitizing enterocolitis (NEC), intraventricular hemorrhagic infarctions (IVH) of Grade III or Grade IV, drug withdrawal, hearing impairment, post hemorrhagic hydrocephalus, and microcephaly.

Table 8

NICU Diagnoses Related to High Probability of Developmental Delay

NICU Diagnoses	Percent of Sample
Birth weight less than 1,200 grams	51.9%
Bronchopulmonary dysplasia	24.1%
Patent ductus arterious	28.9%
Necrotizing enterocolitis	3.6%
Intreventricular hemorrahagic infarctions (IVH) - Grade III or IV	3.6%
Drug exposure	2.4%
Hearing impairment	1.2%
Post hemmorhagic hydrocephalus	4.8%
Microcephaly	1.2%

NICU Follow-Up Clinic and early intervention services. Eighty-four percent of the sample attended the NICU Follow-Up Clinic on the day of the study for a screening appointment with neonatology; the remaining 16% were scheduled for a developmental appointment with developmental pediatrics. At the time of testing, 41 subjects (50.6%) were enrolled in the Maryland Infants and Toddlers program. These subjects were receiving the following services: physical therapy (28.3%, n = 23), occupational therapy (14.8%, n = 12), speech and language therapy (14.8%, n = 12), and special instruction (7.4%, n = 6).

Family characteristics. Most of subjects (59.3%) were accompanied to their routine neonatology or developmental pediatrics NICU Follow-Up appointment by their mother or jointly by their mother and father (27.2%). The majority of caregivers (50.6%) were single; 42% were married, 4.9% were separated or divorced, and 1.2% reported that they were together but not married (1.2% omitted this information). The mothers' mean age was 28.95 years old and the fathers' mean age was 33.74 years old (see Table 9). The parents reported their education in a range from less than high school through Ph.D. Many of the mothers and fathers reported their highest level of education to be a high school degree or GED (33.3% and 32.1%, respectively), with a median level for the mothers at some college and for the fathers at high school.

Geographically, these families resided in 11 of Maryland's 24 counties representative of five Maryland regions: 92.6% resided in Greater Baltimore (consisting of urban, suburban, and rural areas, including Baltimore City), 4.7% lived in the Capital Region (urban and suburban areas), 1.2% lived in Southern Maryland (a rural area), and 2.5% lived in the Eastern Shore (a rural area). Baltimore City, the location of the NICU

Follow-Up Clinic, was the most heavily represented area, reported by 49.4% of the sample.

Table 9

Family Demographics

Demographics	N	%	Median	Mean	Range	SD
Mother's Ages						
14-19	5	6.1%				
20-29	38	46.7%	29 years	28.95	(14-53)	7.24
30-39	32	39.3%				
40-53	5	6.1%				
Information omitted	1	1.2%				
Father's Ages						
18-19	1	1.2%				
20-29	13	16.1%				
30-39	31	38.3%	28 years	33.74	(18-63)	8.06
40-49	6	7.4%				
50-61	2	2.4%				
Information Omitted	28	34.6%				
Mother's Education						
Less than high school	10	12.3%				
High School/GED	27	33.3%				
Some College	21	25.9%				
Associates Degree	8	9.9%	Some College	N/A	N/A	N/A
Bachelor's Degree	9	11.1%				
Master's/Law Degree	3	3.7%				
Ph.D., MD or similar	1	1.2%				
Information omitted	2	2.5%				
Father's Education						
Less than high school	1	1.2%				
High school/GED	26	32.1%				
Some College	6	7.4%	High	N/A	N/A	N/A
Associate's/Vocational	2	18.5%	School			
Bachelor's	10	12.3%				
Master's/Law Degree	5	6.1%				
Information omitted	28	34.6%				
Family Income						
< $10,000	14	17.3%				
$10,000 - $20,000	9	11.1%				
$20,000 - $30,000	10	12.3%	$25,000 -	N/A	N/A	N/A
$30,000 - $40,000	3	3.7%	$30,000			
$40,000 - $50,000	10	12.3%				
>$50,0000	26	32.1%				
Information omitted	9	11.1%				

N/A = not applicable

Validity of the MDS

The goal of the present study was to explore the validity of the MDS when concurrently administered with the BSID-II by examining the instrument's sensitivity, specificity, overall hit rate, and chance hit rate. Following the completion of consent procedures, children were concurrently administered the MDS and BSID-II. The hit rate model was the descriptive method for determining the validity of the MDS. As previously described in greater detail in Chapter III, this type of analysis examines the outcome of a screening measure compared to a child's actual status in terms of sensitivity and specificity, overall hit rate, and chance hit rate. As illustrated in Table 10, the possible outcomes considered were 1) true negative, 2) true positive, 3) false positive, and 4) false negative. A total of 71.6% of the MDS results were classified as true negative; these children passed the MDS and did not demonstrate a developmental delay on the BSID-II. The true positive outcomes of MDS, i.e., the children who failed the MDS and also demonstrated a developmental delay on the BSDI-II, was 18.5%. The false negative percentage was 2.5%; these children passed the MDS but their true developmental status as measured by the BSID-II was greater than 25% delayed. These children represent an error of underprediction of developmental delays. The false positives accounted for 7.4% of the total sample. These children, representing the errors of overprediction of developmental delay, failed according to overall MDS results but their performance on the BSID-II suggests that they were less than 25% developmental delayed.

Table 10

MDS Screening Outcomes

		Criterion (BSID-II) Identified as Delayed (≥25%) (+)	Not Identified as Delayed (<25%) (-)	
Screen (MDS)	Fail Screen (+)	A True Positive 18.5% (n = 15)	C False Positive 28.6% (n = 6)	25.93% (n = 21)
	Pass Screen (-)	B False Negative 2.5% (2)	D True Negative 71.6% (58)	74.07% (n = 60)
		21.90% (n = 17)	79.0% (n = 64)	100% (n = 81)

Sensitivity and Specificity

According to the American Academy of Pediatrics (2001), good developmental screening tests should demonstrate sensitivity and specificity levels of at least 70% and 80% respectively. They report that complexity of assessing the continuous process of child development contributes to the overdetection or underdetection of developmental delays. One of the research questions in this study was asked to determine the concurrent validity of the MDS when compared to the results of the BSID-II when administered to a sample of infants and toddlers born prematurely. The sensitivity, or the percentage of children correctly detected with the MDS as having a developmental delay, was 88.2%. A second research question was asked to determine the specificity of the MDS when concurrently compared to the BSID-II results when administered to a sample of infants

and toddlers born prematurely. The specificity of children correctly who passed the MDS screening due to the absence of a developmental delay was 90.6%. These rates suggest that the MDS is a valid instrument when used to screen developmental status when used with infants and toddlers born prematurely.

Overall and Chance Hit Rates

The final research question sought to determine the overall and chance hit rates of the MDS with the BSID-II when the measures were concurrently administered to a sample of infants and toddlers born prematurely. The overall hit rate, defined as the proportion of accurate screening decisions out of the total number of decisions, was 90.1% (n = 73). This indicates that 90.1% of the 81 MDS administrations yielded a true positive or true negative outcome in terms of the child's actual status. Further exploration of this overall hit rate was necessary because, according to Speece and Cooper (2004), overall hit rates may be "deceptively high when the prevalence of an adverse outcome is low" (p.85). It was therefore important to clarify the rate of subjects who were classified as true negative or true positive on the MDS by chance alone rather than through true classification. This chance hit rate was important because there was an extremely large number of true negatives, or children who passed MDS and passed BSID-II, and therefore a disproportionate large number of hits. The chance hit rate ($p^2 + q^2$, where p = prevalence of condition and q = rate of condition in population) for this sample was determined to be 66.7%. The overall hit rate of 90.1% of the sample population was substantially greater than this chance hit rate. A conclusion that may be drawn is that this 23.4% difference represents an improvement over chance that the MDS will correctly identify children as either in need or not in need of an in-depth evaluation.

Post Hoc Analysis

To eliminate the possibility that SES and other demographic variables contributed to the results, a post hoc analysis was conducted to determine the levels of sensitivity and specificity based upon age of the child, as well as across income levels and race (see Appendix D for Hit Rates). Although the findings of good sensitivity and specificity for the total sample of 81 subjects demonstrated the validity of this instrument with a population of infants and toddlers born prematurely, the decrease in the sample size when various subgroups were analyzed led to some instability of results. Table 11 presents the findings.

Age-based sensitivity and specificity of the MDS. When the age of the child was collapsed into three categories based on year of life, the specificity of the MDS remained high; 100% at year one, 83% at year two, and 82% at year three. In this study, however, the low sensitivity rate at year one, 33%, may have been detrimentally influenced by the small number of results (n = 3) used for the calculation of this rate. There is a strong indication of instability of the results due to the low number of subjects in this subgroup classified as true positive or false negative test. This indicates a need for more research with larger sample sizes.

Table 11

Post Hoc Sensitivity and Specificity

Status	N	Percent of Total Sample	Sensitivity	Specificity
Total Population	81	100%	88.2%	90.6%
Subject Age Visits				
Year 1 (3, 6, & 9 month visits)	32	41%	33%	100%
Year 2 (12, 18, & 24 month visits)	29	37%	100%	83%
Year 3 (30 & 36 months)	20	25%	100%	82%
Yearly Household Income				
Information not provided	9	11%	100%	71%
Less than $25,000	27	33%	89%	95%
$25,000 - $50,000	19	23%	75%	93%
Greater than $50,000	26	31%	100%	92%
Ethnicity				
African American	54	67%	92%	95%
Caucasian	23	23%	75%	89%
Hispanic	1	1%	*	*
Other	1	1%	*	*
Bi-Racial	2	2%	*	*
Mother's Education				
Information not provided	2	4%	100%	100%
Less than high school	10	12%	33%	100%
High school/GED/vocational	27	33%	100%	85%
Some college – Associate's	29	36%	100%	92%
Bachelor's or higher	13	16%	100%	92%
Father's Education				
Information not provided	28	35%	90%	94%
Less than high school	1	1%	*	*
High school/GED/vocational	26	32%	67%	91%
Some college – Associate's	11	14%	100%	90%
Bachelor's or higher	15	19%	100%	85%

* Insufficient data

Income-based sensitivity and specificity of the MDS. The sensitivity and specificity levels were additionally examined across income levels to determine if a person's SES impacted the validity of test results. Thirty-three percent of the subjects' families reported a combined annual income level of less than $25,000. The sensitivity and specificity for this group remained high at 89% and 95% respectively. Of the 23% of families who reported their income to be between $25,000 and $50,000, the sensitivity and specificity was 75% and 93%. Although the sensitivity level in this group decreased, it remained greater than the recommended level of 70% (American Academy of Pediatrics, 2001). Finally, 31% of the sample reported a combined family income of greater than $50,000; sensitivity for this group was determined to be 100% and specificity was 92%. Eleven percent of the total sample did not provide information regarding their income status. These finding suggest that the MDS may have validity when administered to infants and toddlers born prematurely across all income groups, but caution again should be taken in interpretation due to the small number in subjects in each income subgroup.

Race-based Sensitivity and Specificity. The majority of the sample population (67%) was African American and Caucasian (28%). Sensitivity and specificity were calculated for these two groups only, as the other reported race categories did not have enough subjects to conduct analysis (Hispanic n = 1, Other n = 1, Bi-racial n = 2). The sensitivity and specificity levels again remained high across the two race groups. The African American sample had levels of 92% sensitivity and 95% specificity and the Caucasian group demonstrated sensitivity and specificity levels of 75% and 89%.

Education-based sensitivity and specificity. The subjects' mothers' and fathers' education level was also examined to determine any notable trends. Educational categories were collapsed to analyze the following groups: (a) less than high school, (b) high school or GED, (c) associate's degree, some college, or vocational training, and (d) bachelor's degree or higher (including master's degree, law degree, and Ph.D. degree). Similar to the age category of year one, within the categories of mother's education less than high school and father's education high school or GED, the number of true positives and false negatives were notably low; the total number of subjects in each of these two categories was three. It was noteworthy, however, that the sensitivity for both the mothers and fathers in each group was 100% for the two highest levels of education classification (some college/associate's degree/vocational training and bachelor's degree or higher).

An unexpected finding of this study was the relationship between the false negative (pass MDS/fail BSID-II) and false positive outcomes (fail MDS/pass BSID-II) in relation to developmental domain and age of the child. Post Hoc analysis, including an additional chart review, was conducted to explore these relationships.

False Negative Screening Outcomes

The false negative rate, referring to the children who received a pass on the MDS but demonstrated a delay that was 25% or greater on the BSID-II, was 2.5% (n = 2). It was particularly important to investigate possible explanations for these results because these subjects represented children whose developmental needs could be neglected due to the lack of detection of a possible delay. Both children were tested at the 3-month visit. Difficulties in the gross motor domain were the prevalent concern for each child tested

(see Table 12). The first subject passed the MDS yet demonstrated an atypical gross motor quality. Further investigation of this subject's NICU Follow-Up record indicated that the ultimate need for early intervention services matched the developmental status determined with the BSID-II. The child was referred for a physical therapy assessment from the NICU Follow-Up Clinic staff (due to atypical quality of performance) and subsequently received physical therapy services. Although the second subject performed at a greater than 25% delay on the BSID-II, upon examination of test items many failed items were not appropriate for a child who is unable to sit at four months of age.

Table 12

False Negative Screening Outcomes: Pass MDS/Fail BSID-II

Age Tested	**MDS Quality**	**BSID-II**						**Notes**	**Services**
		MDI	**MDI Status**	**MDI Age**	**PDI**	**PDI Status**	**PDI Age**		
2m.	Atypical gross motor	54	Significant delay	<1 m.	63	Significant Delay	<1 m.	2 Standard deviations from the norm and 25% delayed	Physical therapy
4m.	Typical	93	Mild delay	4 m.	83	Within Normal Limits	3 m.	25% delayed on PDI Unable to sit to complete many PDI gross motor items (not appropriate at 4 m.)	None

False Positive Screening Outcomes

The false positive rate, i.e., the percentage of subjects who failed the MDS but were classified with a less than 25% delay on the BSID- II, was 28.6% (see Table 13). Unlike the children with false negative MDS scores, whose adjusted ages ranged from approximately two months old to four months old, the children with false positive scores were older and ranged between 19 months to 34 months of age. Rather than the gross motor domain, the prevalent fail and suspect scores were found in the language domain (fail = 4, suspect = 1). Although for purposes of analysis in this study the BSID-II results were classified as less than 25% delay and greater or equal to 25% delay, the actual BSID-II classification categories of *within normal limits, mild delay*, and *significant delay* were examined to determine any relationship with the MDS scores. Of the subjects who failed the MDS, four were classified with a mild delay and one was classified with a significant delay on either or both the MDI (n = 4) or the PDI (n = 2) of the BSID-II. Only one subject failed the MDS and was within normal limits on both the MDI and PDI. The subject in this case was 31 months and demonstrated a scatter of skills down to 22 months of age; clinical judgment of the researcher indicated that a speech and language deficit was not captured utilizing the BSID-II.

Further analysis was then conducted to determine the early intervention outcome for the children whose screening demonstrated false negative results. Of the four children who failed the language domain and the one failing the cognitive domain of the MDS and passed the BSID-II, four ultimately received speech and language services as recommended by the clinicians at the NICU Follow-Up Clinic. One of the three children who failed the gross motor domain of the MDS also subsequently received physical

therapy services. These results suggest that the over-identification of language issues may not be as problematic as it first appears. The increased number of test items in an in-depth evaluation should be designed for a child who does not do well on a screening that, by the nature of a screening instrument, should have a limited number of test items.

Table 13

False Positive Screening Outcomes: Fail MDS/Pass BSID-II

Age	Failed MDS		BSID-II						Items Failed	Services
	Quality	Domain	MDI	MDI Status	MDI Age	PDI	PDI Status	PDI Age		
19 m.	Typical	Cognitive	75	Mild delay	16 m.	90	Within Normal Limits	17 m.	MDS: Look for ball, point to 4 body parts	None
25 m.	Typical	Language	70	Mild delay	20 m.	76	Mild Delay	22 m.	MDS: Two step command, variegated jargon MDS & BSID-II: Build train with 2 cubes	Occupational therapy Speech & language therapy
27 m.	Typical	Language Gross motor	88	Within normal limits	25 m.	69	Significant delay	22	< 25% Delay MDS: Engage pretend play, MDS & BSID-II: Repeat number sequence, use word sentences, jump up, up steps alternating feet	Speech & language therapy Physical therapy

Table 13 Continued: False Positive Screening Outcomes: Fail MDS/Pass BSID-II

Age	Failed MDS		BSID-II						Items Failed	Services
	Quality	Domain	MDI	MDI Status	MDI Age	PDI	PDI Status	PDI Age		
31 m.	Atypical	Gross motor	79	Mild delay	25 m.	90	Within Normal Limit	27 m.	MDS & BSID-II: Name six pictures, up steps alternating feet, stops making time MDS: Pronouns Inaccurate* BSID-II: Atypical gross motor quality not captured	Occupational therapy Physical therapy Speech & language therapy
31 m.	Typical	Language Gross motor	81	Mild delay	26 m.	91	Within Normal Limits	28 m.	MDS & BSID-II: Prepositions, go up stairs alternating feet MDS: Name objects by use, pronouns	Speech & language therapy
34 m.	Typical	Language	77	Within normal limits	27 m.	112	WNL	37 m.	Inaccurate* BSID-II – speech delay not captured MDI scatter to 22 m.	None

* Inaccurate = administrator's clinical impression

Children Less Than One Month Corrected

Two subjects, based on an age equivalent score of less than one month, demonstrated a delay of greater than 25% on the BSID-II. When their status, however, was based upon their BSID-II developmental score, one was classified as within normal limits (MDI = 93) and the other as mild delay (PDI = 84). The adjusted age at which these subjects were tested indicated that they were at the bottom limit of the testable age. There was no rule in the BSID-II manual regarding how to address this situation. The following rule was made: if the subject's BSID-II age equivalent was less than one month and the developmental score was less than two standard deviations from the norm (within normal limits or mild delay) a subject was considered to be not delayed. MDI and PDI scores for all subjects are provided in Table

Atypical Development

Atypical development is a criterion that qualifies a child to receive Part C early intervention services in Maryland, yet no published screening tool was found to be able to identify this quality of development. In Maryland, COMAR defines atypical development or behavior as when "an abnormal quality of performance and function in at least one of the five developmental areas interferes with current development, and is likely to result in future developmental delay" (http://www.dsd.state.md.us/comar/10/10.09.40.01.htm). The unique quality of the performance component of the MDS enables administrators to score a child's development as atypical so that qualities that necessitate consultation with an appropriate specialist may occur. The atypical or typical decision made by the administrator is based

upon their professional impression using guidelines that are presented in the MDS manual.

Fourteen subjects were classified with an atypical quality of development on the MDS. Nine of these 14 subjects demonstrated a delay of less than 25% as measured by the BSID-II; seven passed the screen with either an overall score of pass or suspect and one failed the screen (see Table 14). An additional chart review indicated that seven of these subjects received subsequent early intervention services. The six children who demonstrated an atypical quality on the MDS and demonstrated a 25% or greater delay on the BSID-II were already receiving early intervention services. This suggests that the MDS may have the ability to identify children who are in need of in-depth evaluation and eligible for early intervention services due to atypical development concerns that would otherwise be missed because of failure to consider the importance of the quality of child's performance.

Table 14

Atypical MDS Scores and Subsequent Early Intervention Services

		BSID-II < 25% Delay	Received early intervention services	BSID-II ≥ 25% Delay	Received early intervention services
MDS Overall Score	Pass/ Atypical	3	1	1	1
	Suspect/ Atypical	4	4	0	0
	Fail/Atypical	1	1	5	5

CHAPTER V

Discussion

The intent of this study was to explore the concurrent validity of the Maryland Developmental Screen (MDS) by examining the instrument's predictive relationship with the Bayley Scales of Infant Development, Second Edition (BSID-II). To address this issue the following three research questions were asked:

1. What is the sensitivity of the MDS when concurrently administered with the BSID-II?
2. What is the specificity of the MDS when concurrently administered with the BSID-II?
3. What is the overall hit rate and chance hit rate of the MDS when compared to the BSID-II?

These questions were explored by concurrently and independently administering the MDS and the BSID-II to a sample of 81 infants and toddlers born prematurely. Results suggest that when used with children born prematurely, the MDS is a valid developmental screening measure. A discussion of these results is presented in this chapter.

A statement made to the President's Commission on Excellence in Special Education (2002) indicates that there is a lack of appropriate screening tools and that good screening tools need to be developed and validated. Hess and colleagues (2004) support this by arguing that there is a need for alternative screening tools which have better predictive utility and are brief and cost effective. An in-depth analysis of 12

instruments suggested that there was no measure that adequately met the screening needs of infants and toddlers born prematurely. The instruments evaluated included Ages and Stages Questionnaire (ASQ), Healthy Steps "Quick Check Sheets", Behavioral Assessment of Baby's Emotional and Social Style (BABES), Child Development Inventories (CDI), Early Intervention Developmental Profile (EIDP), Brigance Screen, Brigance Infant Toddler Screen (BITS), Diagnostic Inventory for Screening Children (DISC), Battelle Developmental Inventory Screening Test, Bayley Infant Neurodevelopmental Screen (BINS), the Parent's Evaluation of Developmental Status (PEDS), and the Denver Developmental Screen-II. Instruments that appeared to be capable of effectively screening for developmental delays were evaluated for the following criteria: (1) developmental screening of all six developmental domains, (2) use of minimal materials, (3) ease and speed in administration, (4) ability to test young infants with an adjusted age of less than term, (5) hands-on assessment, (6) parent involvement in the assessment process, (7) professional evaluation of the quality of the child's performance, and (8) validity of the instrument. Of the 12 tools reviewed, none met each and every criterion. The Battelle and the Denver II come the closest to meeting these screening criteria; however, the Battelle does not provide enough items at each age interval (Ikle & Wittmer, 1995) and the Denver II's specificity and sensitivity has been challenged by many researchers (Committee on Children with Disabilities, 2001; Glascoe et al., 1992).

Results from this study suggest that the MDS may be capable of providing public health, pediatric, and early intervention communities with a new option for screening the development of children born prematurely. The MDS examines six developmental

domains, uses minimal materials, is easy to administer, is designed to test young infants with an adjusted age of less than term, provides hands-on assessment, encourages parent involvement, provides professional evaluation of the quality of the child's performance, and this study demonstrates its concurrent validity. Despite the fact that the MDS tests children whose age at testing is less than full term, the concurrent validity with children whose adjusted age was less than 16 days old could not be measured because the BSID-II does not have the capability to assess children this young. Future validation of the MDS needs to address the validity at this young age through concurrent testing with appropriate instruments. Unfortunately, no tool currently exists for assessing children whose adjusted age at time of testing is less than term (40 weeks gestation). Also, the criterion that a screening instrument should be quick to administer should be tested. The amount of time needed to administer the MDS was not recorded in this or any other study. Although it is clinically suggested that the MDS is a screening tool that can be administered within ten minutes (personal communication with Hussey-Gardner, April 28, 2005), the timing should be empirically measured.

The results of this study suggest that that the MDS is a screening measure which correlates with a child's classification status on the BSID-II, (i.e., less than 25% delayed and greater than or equal to 25% delayed), which in turn relates to the child's eligibility for early intervention services. The sensitivity (referring to the percentage of subjects correctly detected at screening as having a developmental delay) in this study was 88.2%. The specificity, i.e., the percentage of subjects correctly undetected by the screening due to the absence of a developmental delay, was 90.6%. These numbers indicate that the MDS is a valid screening measure of infants and toddlers born prematurely. The

sensitivity and specificity levels of the MDS that were determined in this study provide preliminary evidence that this instrument demonstrates validity that exceeds many other known screening measures. For additional information refer to Table 2 in Chapter II. The sensitivity of the MDS, 88.2%, exceeded the levels of sensitivity found in research that examined the CDI (Shoemaker, et al., 1993; Montgomery, 1999), Denver II (Glascoe et al., 1992), the Brigance (Glascoe, 2002), BITS (Glascoe, 2002) and the Battelle (Ikle & Wittmer). The specificity, 90.1%, also exceeded validity findings in research conducted on the following instruments: CDI (Montgomery, 1999), Denver II (Glascoe et al, 1992), BITS (Glascoe, 2002); and Battelle (Glascoe & Bryne, 1993). Shoemaker and colleagues' (1993) research on the CDI was the only study found to exceed the MDS' specificity level with a report of a level of 92%.

It is important to include in a discussion of sensitivity and specificity levels the consequential validity related to the impact that a screening instrument may have when incorporated into programs. A tool that is overly sensitive will detect a greater number of children who should receive further in-depth evaluation to determine their actual developmental status. This over-referral may be financially costly and present personnel problems for programs because children who are not delayed and not in need of services may be unnecessarily evaluated. Children who are in need of services and who might otherwise not be recognized as delayed, however, are more likely to be identified. A measure that is more specific will most likely lead a more appropriate number of children being referred for in-depth evaluation and therefore will be more cost-effective for a program. The major caveat to this approach, however, is that a tool that is excessively specific may fail to identify children whose actual classification status is delayed,

potentially leading to a postponement or omission of needed early intervention services. It is the opinion of this researcher that it is more desirable for a screening tool to err on the side of increased sensitivity, as does the MDS, so that all children who are potentially eligible for early intervention services receive them in an appropriate and timely manner.

The prevalence rate, i.e., the percentage of children in the sample who were identified as delayed based on the BSID-II, for this study was 28.6%. The prevalence rate in the general population ranges from 5-10% (Litchenstein & Ireton, 1984); the rate in this study, as expected, was higher. The literature indicates that 33% of children born at a low birth weight may develop neurological problems (Nadeau, Boivin, Tessier, Lefebvre, & Robaey, 2001). The prevalence of disabilities for children born at an extremely low birth weight (less than 1000 grams) has been found to be higher, at rates such as 49% (Vohr et al., 2000) and 47% (Whitfield, 1997). Researchers (Vohr et al., 2000) have used BSID-II Mental Developmental Index (MDI) and Psychomotor Developmental Index (PDI) scores to examine the developmental status of children at 18 and 22 months of age. Others (Whitfield et al., 1997) have examined children at school age using other age appropriate assessments. Neither of these studies used percentage of delay to determine the subjects' developmental status. Although it is unknown, the lower prevalence rate of a delay in the present study compared to other studies may be attributed to the inclusion of children in the sample who were born at higher birth weights. Another possible explanation may be the classification method used in this study, which was indicated by a 25 percent delay on the BSID-II as opposed to a classification based on standard deviation from the norm used in other studies (Whitfield et al., 1997; Vohr et al., 2000). Also, the prevalence rate in this study was perhaps lower

due to the fact that some problems are not identified until an older age when preacademic and academic skills emerge (Whitfield et al., 1997).

According to Litchenstein and Ireton (1984), results from hit-rate analysis are situation-specific. The results from this particular study reflect a decision rule (cutoff point) of a developmental delay of 25% or greater as indicated by the BSID-II. This criteria was selected for the study because it is the percentage of delay required in Maryland and many other states for a child to be eligible for early intervention services. (Shackelford, 2004). It was important that empirical evidence gained from this study be translated into practical utility and as the results of the study results may be applicable to many states that follow the same eligibility guidelines as Maryland. If this decision rule were changed, however, the results and hit rate patterns would most likely vary.

The MDS demonstrated a unique ability to identify children who were in need of in-depth assessment based upon an atypical quality of performance of test items. No other published screening instrument was found to meet this criterion. The administrators at the UMMS NICU Follow-Up Clinic demonstrated reliability in their professional evaluation of the identification of atypical quality of performance. In other settings, however, administrators may not have the clinical ability to identify such problems. This is a potential weakness of the MDS and is currently being addressed in Hussey-Gardner's revision of the manual to include specific indicators of atypical development.

Implications for Research

Prior to this concurrent validity study, the MDS had yet to be subjected to validation procedures. The results of this concurrent validity study provide preliminary

evidence that the MDS is a valid developmental screening tool when administered to infants and toddlers born prematurely. Validity of the MDS, however, has by no means been definitely established. As with any instrument, multiple studies must be conducted with different samples to establish validity and there is need to examine consequential validity of this screening instrument relative to social consequences and social implications of MDS use (Messick, 1995).

The consequential validity of the MDS related to the eligibility criteria for early intervention that is obtained from standardized assessments, should be critically examined. The percentage of developmental delay that a child may demonstrate is often based on assessments such as the BSID-II and is the criterion upon which the acquisition of early intervention services in Maryland is based. Although the percentage of delay varies among states, this is the number one method across the country for identifying children who are eligible for early intervention services (Shackelford, 2004). Many assessment tools, however, including the BSID-II, determine a child's developmental status based upon a score's standard deviation from the mean on a norm-referenced instrument. It is questionable whether an instrument that is designed to classify children based on standard deviation scores may appropriately identify children based upon an age equivalent score (i.e., percentage of delay). The BDID-II manual cautions doing this, indicating that deriving developmental ages from the raw scores on the MDI and PDI may result in an age range and not a specific age (Bayley, 1993). For example, in this study, four children failed the MDS and were less than 25% delayed on the BSID-II. The BSID-II classification status for these children, however, was inconsistent as they were

considered to be mildly delayed (see Table 13 in Chapter IV). Although MDI and PDI scores were not analyzed in this study, the subjects' scores are provided in Appendix E.

Although not used for this study, the BSID-II manual's (Bayley, 1993) classifications of within normal limits, mild delay, or significant delay (see appendix F) provides important insight into the issue of eligibility determination for early intervention. For example, if all children who were considered to be mildly delayed (one standard deviation from the norm) qualified for services, more children would be identified and this increase would be costly to programs. It is recommended that the use of the MDS be examined to determine how a decision to classify children (mildly delayed in this example) as eligible for early intervention services impacts the validity of the screening.

When the sample was subjected to post hoc analysis for SES (by ethnicity, family yearly income, and parents' education) the levels of sensitivity and specificity remained generally high. These results were questionable however, and are perhaps explained by the small number of subjects used in calculations in each SES category (e.g., mother's education less than high school, N = 3). The majority of the sample was African American, which was consistent with the rate of prematurity by race within Maryland (March of Dimes, 2005). There were, however, no American Indian or Asian subjects, and there was little representation of Hispanic children born prematurely.

Replicating this study with a much increased sample size will provide more accurate concurrent validity evidence related to the use of the MDS with infants and toddlers born prematurely when accounting for various SES factors.

Hussey-Gardner's (2003) intent was to identify infants and toddlers whose development may delayed; therefore, the research regarding the validity of the MDS should also be extended beyond infants and toddlers born prematurely. Examples of populations with which this study could be replicated include full term and typically developing children, and children who may be at risk for developmental delay (e.g., Early Head Start population). The MDS is designed so that it may be used in a variety of environments and not only in a clinical setting; the validity should therefore be examined when administered in settings including early childhood centers, at children's homes, and in community settings.

The present study concurrently investigated the outcome of the MDS compared to a child's actual status as determined by the criterion measure, the BSID-II. The BSID-II was selected as the criterion measure due to the general consensus that the BSID-II is the "gold standard" in early childhood assessment (Sattler, 1992). The BSID-II is currently undergoing revision and the Bayley Scales of Infant Development, Third Edition, will be available for use in the fall of 2005. Depending on the modifications that have been made to this assessment, replication of this study using the newer version of the instrument will need to be conducted. Neither the BSID-II nor any other tool that was published at the time of the study has the ability the MDS does to test children at an age less than full term as well as at the older ages represented by the MDS. A future investigation of the MDS should include a concurrent validity study that uses a criterion measure that focuses on children whose age is less than 16 days. Unfortunately, to date, there is no criterion measure to validate this age range. The BSID-II examines the hierarchy of child development through the Psychomotor Development Index (PDI) and

the Mental Development Index (MDI) and does not separate testing items into domains. The BSID-II manual indicates that the cognitive, language, motor, and personal/social domains are represented in the test items of the PDI and MDI. There is no mention of the adaptive domain but many tasks relative to this domain are found within the two indexes. Although there were no false positive or false negative results in this study specific to the adaptive domain, further research should be conducted. The false positive and false negative findings in this study were largely in the gross motor and language domains. It may therefore be important to examine each domain represented in the MDS more closely by concurrently using assessments such as the Peabody Developmental Motor Scales as the criterion measure to further investigate specific domains of development.

Implications for Practice

The importance and need for accurate developmental screening of infants and toddlers is reflected in policies from the American Academy of Pediatrics (Committee on Children with Disabilities, 2001), EPSDT, Early Head Start (45 CFR 1304.20), and Part C of the Individuals with Disabilities Education Improvement Act of 2004. Results from this validation study demonstrate the ability of the MDS to meet this need with one high risk population, children born prematurely, and may lead to a specific impact in promoting the development of infants and young children through an assurance of quality of care and appropriate follow-up services.

It is intended that this evidence-based research will ultimately be transformed into practice through the commercial production of the MDS. Public health, pediatric and early intervention communities will then be able to use the MDS when screening the development of infants and toddlers born prematurely to promote quality of care through

an assurance of appropriate referral to early intervention and follow-up services. Prior to publication more research is needed to validate the psychometrics of the MDS. In addition to concurrent validity, predictive validity is particularly important for screening tests (McLean et al., 1996). Predictive validity is related to the extent to which a screening instrument predicts later a child' development. For example, it is unknown whether the MDS that accurately predicts a child's performance on the BSID-II at two months of age will also predict the outcome of the BSID-II when the same child is 12 months of age. It is recommended that longitudinal studies related to the predictive validity of the MDS be conducted, including, perhaps, research on how the predictive validity of the MDS is related to later acquisition and need of early intervention services.

An examination of construct validity of the MDS should also be considered through convergent and discriminate validity testing via correlation with other screening tests (e.g. Bayley Infant Screener). Tests with good convergent validity demonstrate a high positive correlation with tests that measure the same construct; whereas good discriminate validity has low correlation with tests measuring different constructs (McLean et al., 1996). These types of validity tests would to determine how MDS test items in the areas of development represented (i.e., gross motor, adaptive, etc.) are able to differentiate and isolate different skills that measure a child's development in each of the six domains (McLean et al., 1996). This type of validity may be difficult to measure, however, as no other screening instrument currently meets all recommended criteria for effective screening of infants and toddlers born prematurely.

Reliability studies that examine procedural and scoring reliability and test retest of the MDS should additionally be considered. The interrater reliability related to scoring

the MDS that was completed prior to this study demonstrated a high correlation between administrators with a small sample (n = 12). Additional scoring reliability testing as well as procedural reliability is necessary to determine the extent to which examiners are able to follow the administrative procedures in the MDS manual and to examine the accuracy of scoring (McLean et al., 1996). It is particularly important to determine the administrators' accuracy of in identifying atypical development. Test-retest is also important to determine the extent to which MDS scores of the same children remain stable over time (McLean et al., 1996).

In Maryland, if percentage of delay, rather than standard deviation from the norm continues to be the criterion for eligibility (in addition to atypical development), a matrix approach based on a child's age should be considered. According to Shackelford (2005), some states have designed a matrix to differentiate the amount of delay based upon the age of the child. This approach differentiates the amount of delay needed to qualify for early intervention services based on age of a child. This is because, for example, a six-month-old child with a 25% percent delay is considerably different than a 24-month-old with the same percentage of delay (Shackelford, 2005).

Many states allow early intervention eligibility to be determined through a delay based on either standard deviation from the norm or age equivalent, or a combination of both. It is therefore recommended that the validity of the MDS be studied as it relates to qualification criteria other than the 25% delay used in this study (such as standard deviation or other percentages of delay). Additional research should be specifically examine the MDS to explore its validity with the 25% delay criterion compared to other

states criteria (e.g., 50% delay in Oklahoma, 30% delay in Illinois) to determine the impact that different criteria might have on Maryland's early intervention system.

Limitations

The lack of true randomization of the sample was due to the clinical environment of the NICU Follow-up Clinic. A sequential and stratified recruitment of available subjects was conducted to ensure that the sample was representative of the clinic's population. A pitfall of concurrent validity research is the threat to internal validity resulting from the testing history. In this specific study the measure administered first (either the MDS or the BSID-II) may have impacted the subject's performance of a particular task. For example, a child may not have been able to build a tower of cubes during the administration of the first measure but exposure through testing may have resulted in the child learning and succeeding in the task during the next measure. Data was not collected on the order in which the measures were administered; if replication of this study is conducted, this should be considered. Finally, the use of developmental age equivalent scores rather than actual scores on the BSID-II is a potential limitation of this study. As previously discussed, a 25 percent delay, and not standard deviation as intended with the BSID-II, is a criterion for eligibility for early intervention services in Maryland. This fact drove the decision to use the age equivalent scores. This remains, however, in the opinion of this research, that this was the best way to do this study because the 25% criterion is currently the number one way that eligibility is determined throughout the country, not just in Maryland.

Summary

The MDS is quick and easy to administer, uses minimal materials, uses hands-on assessment, evaluates the quality of performance, tests all six developmental domains, tests children at an age less than term, and uses parent report. This concurrent validation study demonstrates that the instrument has both strong sensitivity (88.2%) and specificity (90.6%). The low chance hit rate (66.68%) indicated that the high overall hit rate (90.1%) was not due to chance alone. This information provides evidence that the MDS, when used with infants and toddlers born prematurely, may be a valid method of screening development. The validity of the MDS should continue to be tested, particularly if the instrument is to be published and use is extended beyond a population of infants and toddlers born prematurely.

APPENDIX A:

MARYLAND DEVELOPMENTAL SCREEN

Maryland Developmental Screen

Developed by Brenda Hussey-Gardner, PhD, MPH

Name:_____ DOB:____/____/____

Gestational Age:____weeks Birth Weight:_____grams

Screen Age	DATE	COGNITIVE		LANGUAGE		SOCIAL-EMOTIONAL		ADAPTIVE		GROSS MOTOR		FINE MOTOR		ESULTS
30-32 Weeks PCA		Eye opening in dim light	ORAN	Exhibits stress to loud noises	ORAN	Limited alertness for eye contact	ORAN	Rooting present Rhythmical nonnutritive suck	ORAN ORAN	Flexion in legs during rest Active flexion of legs when alert	ORAN ORAN	Active flexion of arms when alert	ORAN	P S F - T A Y N
32-36 Weeks PCA		Brief focus on object	ORAN	Responds to sound inconsistently	ORAN	Capacity for gentle social interaction	ORAN	Begins to nipple feed	ORAN	Turns head from midline to side Resistance to passive movements legs	ORAN ORAN	Resistance to passive movements arms	ORAN	P S F - T A Y N
36-40 Weeks PCA		Blinks to light	ORAN	Responds to sound with consistency	ORAN	Gazes at face	ORAN	Nipple feeds almost all	ORAN	Flexed posture Stepping	ORAN ORAN	Placing present	ORAN	P S F - T A Y N
Term		Fixes on ring	ORAN	Cries	ORAN	Soothes when picked up	ORAN	Nipple feeds well	ORAN	Clears head in prone	ORAN	Palmar present	ORAN	P S F - T A Y N
1		Tracks ring 30°	ORAN	Throaty noises	ORAN	Quiets to voice/face	ORAN	Can maintain alertness 1hr	ORAN	Head up in prone	ORAN	Hands fisted	ORAN	P S F - T A Y N
2		Tracks ring horizontally & vertically	ORAN	Single vowel sounds 'ooh, aah'	ORAN	Social smile	ORAN	Can sleep 3-4 hours at night	ORAN	Chest up in prone	ORAN	Hands 50% unfisted	ORAN	P S F - T A Y N
3		Tracks ring in a circle	ORAN	Coos when talked to	ORAN	Anticipatory excitement	ORAN	Rooting reflex inhibited	ORAN	Forearm support in prone	ORAN	Reaches for objects, misses	ORAN	P S F - T A Y N
4		Watches own hands	ORAN	Laughs Turns head to voice	ORAN ORAN	Vocalizes, smiles, reaches more for familiar people	ORAN	Brings hands near mouth	ORAN	Rolls supine to side	ORAN	Hands to midline	ORAN	P S F - T A Y N
5		Secures ring Regards Cheerio	ORAN ORAN	Razz	ORAN	Strong maternal attachment	ORAN	Helps hold bottle	ORAN	Rolls prone to supine or supine to prone Hand support in prone	ORAN ORAN	Picks up cube	ORAN	P S F - T A Y N
6		Looks for dropped spoon	ORAN	Says single syllables Holds arms out to be picked up	ORAN ORAN	Responds to own mirror image	ORAN	Swallows strained or pureed foods	ORAN	Sits tripod Supports weight in standing	ORAN ORAN	Raking	ORAN	P S F - T A Y N
7		Pulls out peg	ORAN	Orients to bell upward/indirect	ORAN	Separation anxiety	ORAN	Takes baby cereal & baby foods well	ORAN	Crawls on belly Sits without support	ORAN ORAN	Transfers Holds 1 cube & takes another	ORAN ORAN	P S F - T A Y N
8		Rings bell	ORAN	Combines syllables	ORAN	Distinguishes own name	ORAN	Bites food voluntarily	ORAN	Creeps on all fours	ORAN	Thumb-finger grasp	ORAN	P S F - T A Y N
9		Removes lid from box	ORAN	Imitates playful sounds	ORAN	Enjoys peek-a-boo games	ORAN	Holds, bites, chews cracker	ORAN	Comes to sit	ORAN	Inferior pincer grasp	ORAN	P S F - T A Y N

Key: Located on back.

MARYLAND DEVELOPMENTAL SCREEN -- Page Two

Scree	DATE	COGNITIVE		LANGUAGE		SOCIAL-EMOTIONAL		ADAPTIVE		GROSS MOTOR		FINE MOTOR		ESULTS
10		Stirs in imitation	ORAN	Dada/mama appropriately	ORAN	Responds with gesture to nursery rhymes	ORAN	Drinks from sippee cup when held	ORAN	Pulls to stand	ORAN	Mature pincer grasp Pokes holes	ORAN ORAN	P S F - T A Y N
11		Solves bear behind cup	ORAN	One word Waves bye-bye	ORAN ORAN	Repeats performance if laughed at	ORAN	Finger feeds part of meal	ORAN	Cruises Stands alone	ORAN ORAN	Puts cube in cup	ORAN	P S F - T A Y N
12		Spontaneous scribble	ORAN	Two words 1-step command with gesture	ORAN ORAN	Gives toy on request	ORAN	Drinks from cup with sippee lid independently	ORAN	Walks with one hand held	ORAN	Builds 2 cube tower Overhand grasp	ORAN ORAN	P S F - T A Y N
15		Looks for 'Where is the ball'	ORAN	3 words Points to indicate wants	ORAN ORAN	Wants to be near adults	ORAN	Takes off hat, shoes, socks	ORAN	Walks quickly without falling	ORAN	Puts 6 pegs in	ORAN	P S F - T A Y N
18		Points to 4 body parts	ORAN	7-9 words Points to 2 pictures	ORAN ORAN	Temper tantrums	ORAN	Uses spoon with little spilling	ORAN	Climbs into adult chair	ORAN	Puts 10 cubes in a cup	ORAN	P S F - T A Y N
21		Attends to story	ORAN	10-20 words Uses veregated jargon	ORAN ORAN	Parallel play	ORAN	Unzips zipper	ORAN	Crawls up & down steps Runs	ORAN ORAN	Builds 3 cube tower	ORAN	P S F - T A Y N
24		4 cube train, no chimney	ORAN	Follows 2-step command 2-word sentences	ORAN ORAN	Toys 'mine'	ORAN	Wipes nose if given a tissue	ORAN	Kicks ball forward Goes up steps marking time	ORAN ORAN	Turns pages of book singly	ORAN	P S F - T A Y N
27		Repeats 2 digits	ORAN	Speech half understandable Follows in/on	ORAN ORAN	Engages in pretend role- playing	ORAN	Undresses with help with buttons	ORAN	Jumps up	ORAN	Imitates vertical & horizontal strokes	ORAN	P S F - T A Y N
30		Names or identifies object by use	ORAN	Names 6 pictures Uses pronouns	ORAN ORAN	Points to self in photo	ORAN	Dries hands	ORAN	Goes up steps alternating feet Walks on tip toes	ORAN ORAN	Imitates circular hand movements	ORAN	P S F - T A Y N
33		Gives/shows 1	ORAN	Uses 3-word sentences Uses verb +ing	ORAN ORAN	Has a friend	ORAN	Dresses with help Toilet trained	ORAN ORAN	Throws ball overhand	ORAN	Builds 9 cube tower	ORAN	P S F - T A Y N
36		Names 1 color	ORAN	250-word vocabulary Uses prepositions	ORAN ORAN	Plays interactive games	ORAN	Dresses independently	ORAN	Rides tricycle Balances on 1 foot 2 seconds	ORAN ORAN	Copies circle	ORAN	P S F - T A Y N
39		Understands top/bottom, up/down	ORAN	Answers yes/no ?s Asks "wh" ?s	ORAN ORAN	Expresses regret	ORAN	Washes hands	ORAN	Broad jumps Goes down steps	ORAN ORAN	Opens & closes lid	ORAN	P S F - T A Y N

KEY: **O** = Observed, **R** = Reported, **A** = Atypical quality **N** = Neither observed nor reported. **P** = Pass: Child receives O or R for all items in age, re-screen at next visit, **S** = Suspect: Child receives N in one or more domains but receives O or R for all items in that domain at preceding age range, re-screen at next visit, if two consecutive suspects refer to developmental team, **F** = Fail: Child receives N in current & preceding age range in one or more domains, refer to developmental team. **T** = Typical quality of performance & no A's marked in row, **A** = Atypical quality of performance with at least one item marked A in row, refer as needed. **Y** = Yes child's performance was as parent expected, **N** = No child's performance was not as parent expected (+ indicates better than expected, -indicates not as good as expected).

APPENDIX B:

CONSENT FORM

RESEARCH CONSENT FORM

The University of Maryland, College Park
The University of Maryland Medical Center

Title of Research Project: A Validation of the Maryland Developmental Screen

Principal Investigator: Brenda Hussey-Gardner, Ph.D., Division of Neonatology, University of Maryland School of Medicine.. Phone: 410-328-8782

Co-Investigators:
Abigail McNinch, M.S., School of Education, UMCP Phone: 410-479-9565
Chris Reiner-Hess, Ph.D., Division of Neonatology, UMMS Phone: 410-247-8799

Explanation of Research Project to Subject:

PURPOSE OF STUDY

You and your child have been asked to participate in this research study because your child was born prematurely and attends the NICU Follow-Up Clinic. The purpose of this study is to examine the validity of the Maryland Developmental Screen (MDS) to see if it accurately identifies children whose development may be delayed and who would benefit from more in-depth evaluation.

PROCEDURES

If you agree for your child to participate in this study you will first be asked to provide the investigators with some demographic information about your child and your family (e.g., child's diagnosis, involvement in the early intervention system, caregiver's education level, etc.). If your child is here today to see a neonatologist, he or she will be assessed by the MDS as part of their routine NICU Follow-Up Clinic care. Your child's development will also be tested using another assessment called the Bayley Scales of Infant Development (BSID-II). If your child has an appointment with the Developmental Team, the MDS and the BSID-II will be given in addition to other scheduled assessments.

Participation in this study will take approximately one hour. It will be determined randomly (like the flip of a coin) your child receives the MDS or the BSID-II first. . You will verbally receive results of all assessments conducted and a written report will be provided if you request one.

RISKS/DISCOMFORTS

There are no known risks or discomforts that may occur as a result of your child participating in this study.

BENEFITS

You and your baby may benefit directly by participation in this study because you may learn new information about your baby's development. The results of this study will also help us determine how accurately the MDS assesses the developmental status of children.

COSTS/COMPENSATION

There is no cost or compensation if you agree to participate in this study.

CONFIDENTIALITY OF RECORDS

Any information learned from this study in which you might be identified will be confidential and disclosed only with your permission. If information learned from this study is published, neither you nor your child will be identified by name. Neither your name, nor the name of your child will be associated with the data in any way, and all of your data will be coded with a number for identification. All data will be stored and locked away in the office of the principal investigator, and only project personnel will have access to the data. By signing this form, however, you allow the research study investigator to make your records available to the University of Maryland Medical System (UMMS) and University of Maryland, College Park (UMCP) Institutional Review Board (IRB) and regulatory agencies as required by law. If information learned from this study is published, you will not be identified by name.

RIGHT TO WITHDRAW

Participation in this study is voluntary. You are not obligated to participate in this research. You are free to withdraw your consent at any time. Refusal to participate will not affect your current or future medical care in any way at the University of Maryland Baltimore, University of Maryland Medical System. You will be told of any significant new findings which develop during the study which may affect your willingness to participate in this study.

OFFER TO ANSWER QUESTIONS AND RESEARCH INJURY NOTIFICATION

The principal investigator, Dr. Brenda Hussey-Gardner, or a colleague of Dr. Hussey-Gardner responsible for this research study, has offered to answer any and all questions regarding your participation in this research study. If you have further questions or in the event of a research related injury, you can contact Dr. Hussey-Gardner at the Department of Neonatology, University of Maryland Medical System at 410-328-8782 or the IRB offices of either UMMS (410-706-5037) or UMCP (310-405-4212).

UNIVERSITY STATEMENT

The University of Maryland at Baltimore is committed to the safety of subjects participating in research. Our policies regarding possible risks are stated below.

RESPONSIBILITY FOR RESEARCH RELATED RISKS

Participating in research may result in an injury, as explained above. If you suffer an injury directly related to your participation in this project, the University of Maryland Baltimore and/or one of its affiliated institutions or health care programs will help you obtain medical treatment for the specific injury and provide referrals to other health care facilities, as appropriate. The University of Maryland Baltimore and/or its affiliated institutions or health care groups will not provide you with financial compensation or reimbursement for the cost of care provided to treat a research-related injury or for other expenses arising from a research-related injury. The institution or group providing medical treatment will charge your insurance carrier, you, or any other party responsible for your treatment costs.

MINIMAL RISK STUDIES

The University is committed to providing subjects of its research all rights due them under State and federal law. You give up none of your level rights by signing this consent form or by participating in the research project. Please call the Institutional Review Board (IRB) if you have any questions about your rights as a research subject.

The research in this consent form has been classified as minimal risk by the University of Maryland Institutional Review Board (IRB), a group of scientists, physicians, and other experts. The Board's membership includes persons who are not affiliated with the University and persons who do not conduct research projects. The Board's decision that the research is minimal risk does not mean that the research is risk-free, however. Generally speaking, you are assuming the risks of research participation, as discussed in the consent form. But, if you are harmed as a result of negligence of a research, you can make a claim for compensation. If you believe that you have been harmed through participation in this research as a result of researcher negligence, you can contact the IRB for more information about claims procedures.

YOUR RIGHTS IN THE EVENT OF STUDY-RELATED INJURY

By signing this Consent Form, you are not giving up any legal rights. If this research project is conducted in a negligent manner and you are injured as a direct result, you may be able to recover the costs of care and other damages from the individual organizations responsible for your injury.

FURTHER INFORMATION AND CLAIMS PROCEDURES

Information regarding research may be obtained from the Institutional Review Board (a group established to protect the rights of research subjects) at the following address /telephone number.

Institutional Review Board
University of Maryland
685 West Baltimore Street
Baltimore, Maryland 21201
(410) 706-5037

If you have a study-related injury, you may contact the Institutional Review Board for information about procedures for making claims against those who may be responsible for injuries due to negligence. In addition, the Institutional Review Board can provide further information about the procedure to make claims for injures not related to negligence.

If you agree to join this study, please sign your name below.

NOT VALID WITHOUT THE IRB STAMP OF CERTIFICATION

_____ _____
Caregiver's Signature Date

____ I have read and understand the information on this form.

____ I have had the information on this form explained to me.

_____ _____
Signature of Investigator Date

_____ _____
Witness to Consent Procedures Date
(Optional unless subject is illiterate or
unable to sign

NOTE: Copies of this Consent Form with original signature must be a) retained on file by the Principle Investigator; and b) given to the subject. A copy must also be deposited in the patient's medical record.

APPENDIX C:

DEMOGRAPHIC QUESTIONNAIRE

Validation of the Maryland Developmental Screen

DEMOGRAPHIC INFORMATION

Child's Code #:

1. **Child's Name:** _____
2. **Sex:** Male Female
3. **Date of Birth:** ___/___/_____
4. **Gestational Age**: _____ **5. Birth weight**: _____
6. **Child's Race:**

   African American Caucasian Hispanic Other_____

7. **Caregiver(s) Present at Exam:**

   Mother Father Grandmother Grandfather Other

8. **Child's Current diagnosis:**

9. **County of Residence:** _____
10. **Type of Childcare Arrangement (if any):** _____
11. **Child's early intervention services:**

Type	Frequency					Location		
PT	1xwk	2xmo	1xmo		other	Home		Center
OT	1xwk	2xmo	1xmo	other		Home	Center	
SPL	1xwk	2xmo	1xmo	other		Home	Center	
SI	1xwk	2xmo	1xmo	other		Home	Center	
SC	1xwk	2xmo	1xmo	other		Home	Center	

12. Parent's marital status: Single Married Separated

13. Parent's ages: Mother _____ Father _____

14. Parent's highest level of education completed:

Mother	**Father**
Less than 9th grade	Less than 9th grade
Some high school	Some high school
GED	GED
High School	High School
Some college	Some college
AA Degree	AA Degree
Bachelor's degree	Bachelor's degree
Master's degree	Master's degree
Ph.D., M.D. or similar degree	Ph.D., M.D. or similar degree

15. Parent's employment status and type of work:

_____

_____

16. Combine household yearly income:

Less than $10,000

Between $15,000 and $20,000

Between $20,000 and $25,000

Between $25,000 and $30,000

Between $30,000 and $40,000

Between $40,000 and $50,000

APPENDIX D:

HIT RATES

HIT RATES

Age-Based Hit Rates

3, 6, 9 Months

MDS	BSID-II	
	>25% (+)	<25% (-)
Fail(+)	1	0
Pass/Suspect (-)	2	29

Sensitivity = 33% Specificity = 100%

12, 18, & 24 Months

MDS	BSID-II	
	>25% (+)	<25% (-)
Fail (+)	11	3
Pass/Suspect (-)	0	15

Sensitivity = 100% Specificity = 83%

30 & 36 Months

MDS	BSID-II	
	>25% (+)	<25% (-)
Fail (+)	3	3
Pass/Suspect (-)	0	14

Sensitivity = 100% Specificity = 82%

Ethnicity-Based Hit Rates

African American

MDS	BSID-II	
	>25% (+)	<25% (-)
Fail (+)	12	3
Pass/Suspect (-)	1	38

Sensitivity = 92% Specificity = 95%

Caucasian

MDS	BSID-II	
	>25% (+)	<25% (-)
Fail (+)	3	2
Pass/Suspect (-)	1	17

Sensitivity = 75% Specificity = 89%

Combined Household Yearly Income

<$25,000

MDS BSID-II

	>25% (+)	<25% (-)
Fail (+)	8	1
Pass/Suspect (-)	1	17

Sensitivity = 89% Specificity = 95%

> $25,000 < $50,000

MDS BSID-II

	>25% (+)	<25% (-)
Fail (+)	3	1
Pass/Suspect (-)	1	14

Sensitivity = 75% Specificity = 93%

> $50,000

MDS BSID-II

	>25% (+)	<25% (-)
Fail (+)	2	2
Pass/Suspect (-)	0	22

Sensitivity = 100% Specificity = 92%

Mother's Education Hit Rates

< High School

MDS BSID-II

	>25% (+)	<25% (-)
Fail (+)	1	0
Pass/Suspect (-)	2	7

Sensitivity = 33% Specificity = 100%

High School/GED/Vocational

MDS BSID-II

	>25% (+)	<25% (-)
Fail (+)	7	3
Pass/Suspect (-)	0	17

Sensitivity = 100% Specificity = 85%

Some College/Associate's Degree

MDS BSID-II

	>25% (+)	<25% (-)
Fail (+)	5	2
Pass/Suspect (-)	0	22

Sensitivity = 100% Specificity = 92%

> Bachelor's Degree

MDS BSID-II

	>25% (+)	<25% (-)
Fail (+)	1	1
Pass/Suspect (-)	0	11

Sensitivity = 100% Specificity = 92%

Father's Education Hit Rates

< High School

MDS BSID-II

	>25% (+)	<25% (-)
Fail (+)	0	0
Pass/Suspect (-)	0	1

Insufficient cell sizes for sensitivity and specificity = %

High School/GED/Vocational

MDS BSID-II

	>25% (+)	<25% (-)
Fail (+)	2	3
Pass/Suspect (-)	1	23

Sensitivity = 67% Specificity = 91%

Some College/Associate's Degree

MDS BSID-II

	>25% (+)	<25% (-)
Fail (+)	1	0
Pass/Suspect (-)	0	7

Sensitivity =100% Specificity = 90 %

> Bachelor's Degree

MDS BSID-II

	>25% (+)	<25% (-)
Fail (+)	2	2
Pass/Suspect (-)	0	11

Sensitivity = 100% Specificity = 85%

APPENDIX E:

BSID-II MDI & PDI SCORES BY VISIT MONTH

Mental Developmental Index (MDI) Scores by Age Visit

MDI	Visit Month								Total
	3	6	9	12	18	24	30	36	
<50						1			1
50			1		1	1			3
52								1	1
54		1							1
55					1				1
56					1				1
58					2				2
62					1	1	1		3
65								1	1
66						1			1
70						1		1	2
72						1			1
75					2			1	3
77							1	1	2
78							1		1
79							1		1
81							1		1
82							1		1
84	1		1	2	1				5
85		1						1	2
87							1		1
88			1	2	1	1			5
89			1				2		3
90	2			1				2	5
91			1						1
92			1			1			2
93		1							1
94								1	1
95		1							1
96		3	2						5

MDI Scores by Age Visit, continued

MDI	Visit Month								Total
	3	6	9	12	18	24	30	36	
97			1						1
98	2		2	1			1		6
100	2			1					3
101			1						1
102	1								1
103		1		1					2
104	1								1
105		1							1
106	1								1
111				1					1
114				1					1
122			1						1
132						1			1
142							1		1
Total	**10**	**9**	**13**	**10**	**10**	**9**	**11**	**9**	**81**

Psychomotor Developmental Index (PDI) Scores by Visit Month

PDI	Visit Month								Total
	3	6	9	12	18	24	30	36	
45				1					1
50								1	1
52					1				1
55					1				1
57						1			1
59					1				1
63		1							1
66			1						1
67					1		1		2
69						1			1
72			1						1
76			2			1			3
77						1			1
78					1			1	2
79								1	1
80		1							1
81			1						1
83		1		1					2
84	1								1
85		1				2			3
87					1				1
88			1				1		2
89				1			2		3
90	1				2		1		4
91			1	2			1		4
92	2								2
93	1							1	2
94		1	2		1		1		5
95			1				1	1	3
96	1								1
97				1		1		1	3

PDI Scores by Visit Month, continued

PDI	Visit Month								T
	3	6	9	12	18	24	30	36	
98								1	1
99						1			1
100		1	1		1	1			4
101				1					1
102		2							2
103							1		1
104				2					2
105		1		1					2
107	1								1
108			1					1	2
110							1		1
112								1	1
113	3		1						4
114							1		1
Total	**10**	**9**	**13**	**10**	**10**	**9**	**11**	**9**	**81**

APPENDIX F:

BSID-II MDI & MPI CLASSIFICATION STATUS

BSID-II Classification for All Children

		PDI			Total
		Within Normal Limits	Mild Delay	Significant Delay	
MDI	Within Normal Limits	49	7	0	56
	Mild Delay	7	3	0	10
	Significant Delay	6	3	6	15
	Total	62	13	6	81

BSID-II Classification for Children Attending 3, 6, and 9 Month Visit

		PDI			Total
		Within Normal Limits	Mild Delay	Significant Delay	
MDI	Within Normal Limits	25	3	0	28
	Mild Delay	1	1	0	2
	Significant Delay	0	1	1	2
	Total	26	5	1	32

BSID-II Classification for Children Attending 12, 18, and 24 Month Visit

		PDI			Total
		Within Normal Limits	Mild Delay	Significant Delay	
MDI	Within Normal Limits	13	2	0	15
	Mild Delay	2	2	0	4
	Significant Delay	6	1	3	10
	Total	21	5	3	29

BSID-II Classification for Children Attending 30 and 36 Month Visit

		PDI			Total
		Within Normal Limits	Mild Delay	Significant Delay	
MDI	Within Normal Limits	11	2	0	13
	Mild Delay	4	0	0	4
	Significant Delay	0	1	2	3
	Total	15	3	2	20

REFERENCES

Als, H., Lester, B.M., Tronick, E.Z., & Brazelton, T.B. (1982). Toward a research instrument for the assessment of preterm infants' behavior (APIB). In H. Fitzgerald, B.M. Lester, & M.W. Yogman (Eds*), Theory and research in behavioral pedatrics, (1),* (pp. 35-132). New York: Plenum.

Als, H. (1997). Earliest intervention for preterm infants in the Newborn Intensive Care Unit. In Michael Guralnick (Ed.). *The Effectiveness of Early Intervention,* pp. 47-76.

Aylward, G.P. (1995). *Bayley Infant Neurodevelopental Screener*. San Antonio, TX: Psychological Corporation.

Barrera, M., Kitching, K., Cunningham, C., Doucet, D., & Rosenbaum, P. (1991). A three-year early home intervention follow-up study with low birth weight infants and their parents. *Topics in Early Childhood Special Education, 10,* (4), 14-28.

Bernbaum, J. C. & Batshaw, M. L (1997). Ch 7. Born too soon, born too small. In *Children with disabilities: A medical primer.* (Eds.). Baltimore: Paul H. Brooks.

Bayley, N. (1993*). Bayley Scales of Infant Development, 2nd. Ed*. The Psychological Corporation.

Beal, J.A., Tiani, T.B., Saia, T.A., & Rothstein, E.E. (1999). The role of the neonatal nurse practitioner in post NICU follow-up. *Journal of Perinatal Neonatal Nursing, 13, (1),* 78-89.

Benish, J. K. (1998). Review of the Bayley Infant Neurodevelopmental Screener, In J. C. Impara & B. S. Plake (Ed.), *The Thirteenth Mental Measurements Yearbook*, (pp.86-88). Lincoln, Nebraska: The University of Nebraska Press.

Bennet, F. & Scott, D. (1997). Long-term perspective on premature infant outcome and contemporary intervention issues. *Seminars in Perinatology, 21,* (3), 190-201.

Berger, S.P., Holt-Turner, I., Cuppoli, J.M, Mass, M., & Hagermen, J. (1998). Caring for the graduate from the neonatal intensive care unit. *Pediatric Clinics of North America, 45,* (3), 701-7012.

Brazelton, T. B. (1973). *Neonatal Behavioral Assessment Scale*. (Clinics in Developmental Medicine No. 50). Philadelphia: Lippincott.

Brazelton, T. B. (1994). Behavioral Competence. In G. B. Avery, M. A. Fletcher, & M. G. MacDonald (Ed.) *Neonatology: Pathophysiology and management of the newborn, fourth edition,* (pp. 289-300). Philadelphia: J. B. Lippincott Company.

Brigance, A. (1990). *Brigance Preschool Screen*. North Billerica, MA: Curriculum Associates, Inc.

Brigance A. & Glascoe, F.P. *Brigance Infant and Toddler Screens*. North Billerica, MA: Curriculum Associates, Inc.

Bull, M., Bryson, C., Schreiner, R., & Lemons, J. (1986). Follow-up of infants after intensive care. *Perinatology Neonatology, 10,* 23-38.

Carey, W.B. (2002). Rapid, competent, and inexpensive developmental-behavioral screening is possible. *Pediatrics, 109,* (2), 316-318.

Censullo, M. (1994). Developmental delay in healthy premature infants at age two years: implications for early intervention. *Developmental and Behavioral Pediatrics, 15,* (2), 99-104.

Committee on Children with Disabilities (2002). Developmental surveillance and screening of infants and young children. *Pediatrics, 108,* (1), 192-195.

Doig, K.B., Macias, M.M., Conway F.S., Craver, J.R., & Ingram P.E. (1999). The Child Development Inventory: A developmental outcome measure for follow-up of the high-risk infant. *The Journal of Pediatrics, 135,* 35-362.

Dobrez D., Sasso A.L., Holl J., Shalowitz M., Leon S., & Budetti P. (2001). Estimating the cost of developmental and behavioral screening of preschool children in general pediatric practice. *Pediatrics, 108,* (4), 913-922.

Frankenburg, W. K. (2002). Developmental surveillance and screening of infants and young children. *Pediatrics, 109,* (1), 144-146.

Frankenburg, W.K., Dodds, J., Archer, P., Bresnick, B., Maschka, P., Edelman, N., & Shapiro, H. (1992). *Denver II, 2nd Training Manual, 2nd Edition*. Denver Colorado: Denver Developmental Materials, Incorporated.

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.

Glascoe, F.P. (2002). The Brigance Infant and Toddler Screen: Standardization and validation. *Journal of Developmental Pediatrics, 23,* (3), 145-150.

Glascoe, F.P. & Byrne, K.E. (1993). The usefulness of the Battelle Developmental Inventory Screening Test. *Clinical Pediatrics, 32*, (5), 273-280.

Glascoe, F.P., Byrne, K., Ashford, L.G., Johnson, K.L., Chang, B., & Strickland, B. (1992). Accuracy of the Denver-II in developmental Screening. *Pediatrics, 89,* (6), 1221-1225.

Hack, M. & Sanaroff, A. (1999). Outcomes of children of extremely low birth weight and gestational age in the 1990's. *Early Human Development, 53,* 193-218.

Hall, R.T. (2000). Prevention of premature birth: Do pediatricians have a role? *Pediatrics, 105*, (5), 1137-1140.

Harris, S.R. & Daniels, L.E. (2001). Reliability and validity of the Harris Infant Neuromotor Test. *The Journal of Pediatrics, 139*, 249-253.

Hussey-Gardner, B. (2003). Unpublished Manuscript, University of Maryland, *Maryland Developmental Screen – User's Manual*. Baltimore.

Hess CR, Kerr MA, Black MM: Use of the Bayley Infant Neurodevelopmental Screener with an Environmental Risk Group. *Journal of Pediatric Psychology, 29*, (5): pages 321

-341, 2004.

Humphries-Wadsworth, T. (1998). *Emerging/evolving views of the meaning of score validity*. Paper presented at the Southwestern Psychological Association, New Oreals, LA (ERIC Document Reproduction Service No. ED418129)

Hussey-Gardner, B.T., Wachtel, R.C., & Viscardi, R.M. (1998). Parent perceptions of an NICU Follow-Up Clinic. *Neonatal Network, 17*, 33-39.

Hussey-Gardner, B. (1995). *Evaluation of family-centered information and instruction sessions in a neonatal intensive care unit follow-up clinic.* Unpublished doctoral dissertation, The University of Maryland, College Park.

Hussey-Gardner, B., McNinch, A., Anastasi, J., & Miller, M. (2003). Early intervention best practice: Collaboration among a NICU, and early intervention program, and a NICU Follow-Up Program. *Neonatal Network, 21,* (3), 15-22.

Ikle, L. & Wittmer, D. (1995). Screening instruments: Review of instruments for screening children ages birth to five years. Updated. Denver: Colorado State Department of Education. Ireton, H. (1992). *Child Development Inventory Manual*. Minnesota: Behavior Science Systems, Inc.

Isaac, S. & Michael, W.B. (1995). *Handbook in research and evaluation: For education and the behavioral sciences, 3rd. ed.* San Diego: Educational and industrial testing services.

Korner, A.F., Stevenson, D.K., Spiker, D., Scott, D., Constantinou, J., & Dimiceli, S. (1993). Prediction of the development of low birth weight preterm infants by a new neonatal medical index. *Developmental and Behavioral Pediatrics, 14*, (2), 106-111.

Krug, D. (1998). Review of the Bayley Infant Developmental Screener. In J. C. Impara & B. S. Plake (Ed.), *The Thirteenth Mental Measurements Yearbook,* (pp. 88-89). Lincoln, Nebraska: The University of Nebraska Press.

Kube, D.A., Wilson W.M., Petersen, M.C., & Palmer, F.B. (2000). CAT/CLAMS: its use in detecting early childhood cognitive impairment. *Pediatric Neurology, 23*, (3), 208-215.

LaRossa, M. M. (2000). Understanding preterm infant behavior in the NICU. Retrieved March 1, 2005 from httpp://www.pediatrics.emory.edu/NEONATOLOGY/DPC/nicubeh.htm

Leonard, C.H., Piecuch, R.E., & Cooper, B. A. (2001). Use of the Bayley Infant Neurodevelopmental Screener with Low Birth Weight Infants. *Journal of Pediatric Psychology, 26,* (1), 33-40.

Lichtenstein, R. & Ireton, H. (1984). *Preschool screening. Identifying children with developmental and educational Problems.* Orlando, Florida: Grune & Stratton, Inc.

Linden D., Paroli, E., & Doran, M. (2000). *Preemies: The essential guide for parents of premature babies.* New York, New York: Pocket Book, 49-52.

Macias, M. M., Saylor, C. F., Greer, M. K., Charles, J. M., Bell, N., & Katikaneni, L. D. (1998). Infant screening: The usefulness of the Bayley Infant Neurodevelopmental Screener and the Clinical Adaptive Test/Clinical Linguistic Auditory Milestone Scale. *Developmental and Behavioral Pediatrics, 19,* (3), 155-160.

Mahoney, G., Kaiser, A., Girolametto, L., MacDonald, J., Robinson, C., Stafford, P., et al. (1999). Parent education in early intervention: A call for a renewed focus. *Topics in Early Childhood Special Education, 19,* (3), 131-140.

Maryland Infants and Toddlers Program (2000). *Dreams & challenges: A family's guide to the Maryland Infants & Toddlers Program.* Maryland Infants and Toddlers Program Maryland State Department of Education.

March of Dimes Birth Defects Foundation (2003). Born too soon and too small in the United States. Retrieved January 18, 2005 from http://www.marchofdimes.com/peristats/prematurity.aspx?=99

Matilainen, R. (1987). The value of correction for age in the assessment of prematurely born children. *Early Human Development, 15*, (5), 257-64.

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.

Meisels, S. J. & Provence, S. (1989). *Screening and assessment: Guidelines for identifying young disabled and developmentally vulnerable children and their families*. Washington, GC: National Center for Clinical Infant Programs.

Messick, S. (1995). Validity of psychological assessment: validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, (90), 741-749.

Messick, S. (1995). Meaning and values in test validation: the science and ethics of assessment. *Educational Researcher, 18*, (2), 5-11.

McCarton, M., Brooks-Gunn, J., Wallace, I., Bauer, C., Bennett, F., Bernbaum, J., et al., (1996). Results at age 8 years of early intervention for low-birth weight premature infants. *Journal of the American Medical Association, 277* (2), 126-132.

Mcgrath, M. M., Sullivan, M.C., Lester, B.M., & Oh, W. (2000). Longitudinal neurologic follow-up in neonatal intensive care unit survivors with various neonatal morbidities. *Pediatrics, 106*, (6), 1397-13405.

McLean, M., Bailey, D.B., & Wolery, M. (1996*). Assessing infants and preschoolers with special needs*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Montgomery, L.A. (1999). Making a multidisciplinary neonatal developmental care team a reality. *Neonatal Network, 18,* (4), 47-49.

Nadeau, L., Boivin, M., Tessier, R., Lefebvre, F., & Robaey, P. (2001). Mediators of behavioral problems in 7-year-old children born after 24 to 28 weeks of gestation. *Journal of Developmental and Behavioral Pediatrics, 22,* (1), 1-10.

Neutens, J.J. & Rubinson L. (1997). *Research techniques for the health sciences, second edition*. Needham Heights, Massachusetts: Allyn & Bacon.

Palisano, R.J. (1986). Concurrent and predictive validities of the Bayley Motor Scale and the Peabody Developmental Motor Scales. *Physical Therapy, 66,* 1714-1719.

Perlman, J.M. (2001). Neurobehavioral deficits in premature graduates of intensive care--potential medical and neonatal environmental risk factors. (Review Article). *Pediatrics, 108,* 1339-1348.

Provost, B., Crowe, T.K., & McClain, C. (2000). Concurrent validity of the Bayley Scales of Infant Development II Motor Scales and the Peabody Developmental Motor Scales in two-year-old children. *Physical Occupational Therapy Pediatrics, 20,* 5-18.

Resnick M.B., Eyler, F.D., Nelson, R.M., Eitzman, D.V., & Bucciarelli, R. L. (1997). Developmental intervention for low birth weight infants: Improved early developmental outcome. *Pediatrics, 80,* 68-74.

Rickards, A. L., Kitchen, W. H., Doyle, L. W., & Kelly, E. A. (1989).

Correction of development and intelligence test scores for premature birth. *Australian Paediatric Journal, 25*, (3), 127-129.

Rogers, S.J., Donovan, C. M., D'Eugenio, D., Brown, S. L., Whiteside Lynch, E., et al., 1981). *Early Intervention Development Profile*. University of Michigan Press: Ann Arbor, MI.

Rosenbach, M.L. & Gavin, N.I. (1998). Early and Periodic Screening, Diagnosis, and Treatment and managed care. *Annual Reviews of Public Health, 19,* 507-525.

Rossman, M.J., Hyman, S.L., Rorabaugh, M.L., Berlin, L.E., Allen, M.C., & Modlin, J.F. (1994). The CAT/CLAMS assessment for early intervention services: Clinical adaptive test/clinical linguistic auditory milestone scale. *Clinical Pediatrics, 33,* 404-409.

Saigal, S., Rosenbaum, P.L., Feeny, D., Burrows, E., Furlong, W., Staskopf, B.L., Hoult, L. (2000). Parental perspectives of the health status and health-related quality of life of teen-aged children who were extremely low birth weight and term controls. *Pediatrics, 105,* 569-574.

Salvia, J. & Ysseldyke, J.E. (1999). *Assessment, eighth edition*. Houghton Mifflin Company: Boston.

Sattler, J.M. (1992). *Assessment of children: revised and updated third edition*. San Diego: Jerome M. Sattler, Publisher, Inc.

Singer, L.T., Siegal, C., Lewis, B., Hawkins, S., Yamashita, T., & Baley, J. (2001). Preschool Language Outcomes of Children with History of

Bronchopulmonary Dysplasia and Very Low Birth Weight. *Journal of Developmental & Behavioral Pediatrics, 22,* (1), 19-26.

Schwarting, G. (1998). Review of the Diagnostic Inventory for Screening Children. In J. C. Impara & B. S. Plake (Ed.), *The thirteenth mental measurements yearbook,* (pp. 365-366). Lincoln, Nebraska: The University of Nebraska Press.

Shackelford, J. (2004). *State and jurisdictional eligibility definitions for infants and toddlers with disabilities under IDEA.* (NECTAC Notes No. 16.) Chapel Hill: The University of North Carolina. FFG Child Development Institute. National Early Childhood Technical Assistance Center.

Shoemaker, O.S., Saylor, C.F., & Erickson, M.T. (1993). Concurrent validity of the Minnesota Child Development Inventory with high risk infants. *Journal of Pediatric Psychology, 18,* (3), 377-388.

Sonnader, K. (2000). Early identification of children with developmental disabilities. *Acta. Paediatr. Suppl., 433,* 17-23.

Speece, D.L. & Cooper, D.H. (2004). Methodological issues in research on language and early literacy from the perspective of early identification and instruction. In B. Shulman, K. Apel, B. Ehren, E.R. Silliman & Stone (Eds.), *Handbook on Language and Literacy Development and Disorders*. New York: Guilford Press.

Squires, J., Bricker, D., & Potter, L. (1997). Revision of a parent-completed development screening tool: Ages and Stages Questionnaires. *Journal of Pediatric Psychology, 22,* (3), 313-328.

Subramanian S. K.N., Yoon H., & Toral, J.C. (2002). Extremely low birth weight infant. Emedicine Journal, 10, (3), http://author.emedicine.com/ped/topic2784.htm.

Turnbull, R., Turnbull, A., & Wheat, M. (1982). Assumptions about parental participation: a legislative history. *Exceptional Education Quarterly*, 1-7.

Wachtel, R.C., Sapiro, B.K., Palmer, F.B., Allen, & Capute, A.J. (1994). CAT/CLAMS. A tool for the pediatric evaluation of infants and young children with developmental delay. Clinical Adaptive Test/Clinical Linguistic and Auditory Milestone Scale. *Clinical Pediatrics, 33,* 410-415.

Watson, S. & Henington, C. (1998). Review of the Diagnostic Inventory for Screening Children. In J. C. Impara & B. S. Plake (Ed.), *The thirteenth mental measurements yearbook*, (pp. 366-368). Lincoln, Nebraska: The University of Nebraska Press.

Weisglas-Kuperus, N., Baerts, W., Smrkovsky, M., & Sauer, P. (1993). Effects of biological and social factors on the cognitive development of very low birth weight children. *Pediatrics, 92,* (5), 658-655.

Wilson, S. L. & Michaeleen Cradock, M. (2004). Review: Accounting for prematurity in developmental assessment and the use of age-adjusted scores. *Journal of Pediatric Psychology, 29,*(8), 641-649.

Woods, N.S., Marlow, N., Costeloe, K. (2000). *Neurologic and developmental disability after extremely preterm birth.* ISSER, 154, 725-731.