Concurrent Validity of the Transdisciplinary Play Based Assessment-2

Keri E. Linas
University of Denver

Follow this and additional works at: https://digitalcommons.du.edu/etd

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
https://digitalcommons.du.edu/etd/369
Concurrent Validity of the Transdisciplinary Play Based Assessment

———

A Dissertation

Presented to

the Morgridge College of Education

University of Denver

———

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

———

by

Keri E. Linas

August 2009

Advisor: Toni Linder, Ed.D
ABSTRACT

Standardized assessment tools have been criticized for use with young children; however, there is limited empirical research supporting developmentally appropriate, play-based assessments. Forty children, with and without developmental disabilities, ages 10-73 months referred from the community of a large city received a transdisciplinary play based assessment (TPBA-2) and standardized assessment. Caretakers also completed the Vineland Scales of Adaptive Behavior. Concurrent validity between the TPBA-2 and standardized assessment tools (Bayley Scales of Infant and Toddler Development-III and the Mullen Scales of Early Learning) was demonstrated (p < .01). Ecological validity of the TPBA-2, or the ability of the assessment tool to capture every day functional skills, was also demonstrated (p < .01). No differences were found between the type of assessment and the facilitation of the child’s typical skills observed in the home and community. No differences were found between the type of assessment received and the child’s behavior during the assessment. This study provides empirical support of the TPBA-2 as a developmentally appropriate assessment tool in compliance with the standards of the Individuals with Disabilities Education Improvement Act (IDEIA).
Acknowledgements

I would like to thank Toni Linder for her energy, commitment, and absolute passion for early childhood development, assessment, and education as well as her dedication to the training of students. Her creativity, flexibility, and exquisite ability to think beyond convention allowed me to pursue the joint doctoral degree.

Thank you Lavita Nadkarni, Gloria Miller, and Fernand Lubuguin for your additional support and guidance with the dissertation process. Lavita and Fernand are responsible for shaping my learning experience and professional development as an ethical and culturally competent psychologist. I have appreciated their patience and respect for individual development as well as their kind, constructive, and affirming mentorship. I am grateful to Gloria’s flexibility and support towards the pursuit of the joint doctoral degree.

A special thank you to Brittney Bixby, Staci Jordan, and Laura Kavlie for their hard work, attention to detail, organization and professionalism and whose efforts were crucial in the data collection for the study. Thank you also to Maura Ryan whose efforts helped to finalize the data collection.

Finally, thank you to all of the parents, children, and professionals who volunteered their time and made this research study possible.
# Table of Contents

Chapter One: Introduction

Historical Approaches to Assessment ................................................................. 5
Limitations to Historical Approaches ................................................................. 7
Best Practices for the Assessment of Young Children ....................................... 25
Validity .............................................................................................................. 37
Purpose of the Current Study ............................................................................ 45

Chapter Two: Method...................................................................................... 49
Materials for Measurement ............................................................................. 51
Procedure ......................................................................................................... 63

Chapter Three: Results .................................................................................. 69
Preliminary Results ......................................................................................... 69
Secondary Results ............................................................................................ 74

Chapter Four: Discussion .............................................................................. 86
Other Findings .................................................................................................. 98
Limitations ....................................................................................................... 100
Future Directions ............................................................................................. 102
Clinical Implications ....................................................................................... 104

References ....................................................................................................... 108

Appendices ...................................................................................................... 120
Appendix A ...................................................................................................... 120
Appendix B ...................................................................................................... 128
Appendix C ...................................................................................................... 132
Appendix D ...................................................................................................... 138
Appendix E ...................................................................................................... 141
Appendix F ...................................................................................................... 146
Appendix G ...................................................................................................... 147
Appendix H ...................................................................................................... 149
Appendix I ...................................................................................................... 153
CHAPTER 1

Introduction

The Individuals with Disabilities Education Act (IDEA) (1990, 1997) and the Individuals with Disabilities Education Improvement Act (IDEIA) (2004) established a federal mandate for the provision of educational and developmental services to children (U.S Department of Education, n.d). The authority of IDEIA is experienced at the federal, state, and community levels because it shapes how states and public agencies provide assessment, intervention, special education, and related services to infants, toddlers, preschool children, and youth. The 2004 Individuals with Disabilities Improvement Act defined a child with a disability as:

A child with mental retardation, hearing impairments (including deafness), speech or language impairments, visual impairments (including blindness), serious emotional disturbance …., orthopedic impairments, autism, traumatic brain injury, other health impairments, or specific learning disabilities; and developmental and emotional disabilities (U.S. Department of Education, n.d, p.2).

Identifying children eligible for intervention involves a complex system of identification, referral, and assessment (Dunst & Trivett, 2004). Assessment is a broad term (McLean, 2004) that refers to formally collecting information about a child’s skills and behaviors. IDEA outlines three criteria for choosing the tools for assessment of young children. First, the tool for assessment should be based on informed clinical
opinion. Second, tools or tests should demonstrate reliability and validity for the purposes for which they are used. Third, the assessment process should be multisourced, with unbiased tests and procedures, completed by a multidisciplinary team (Fewell, 1991; Harrington & Tongier, 1993). The goals of assessment in early childhood are to gather information about the child by eliciting the most representative or typical sample of the child’s behavior to inform diagnosis and treatment goals as well as to gather information about the strengths or functional abilities of the family (Calhoon, 1997; Mesiels & Atkins-Burnett, 2001).

There continues to be a question about which measures can successfully elicit the most typical range of skills and behaviors for the individual young child as well as gather information about the family (Bracken, 1994, Calhoon, 1997; Neisworth & Bagnato, 2000). In older children, traditional standardized, norm-referenced tools are considered to be the most effective and efficient. However with the passage of Public Law 99-457 in 1986 (P.L: 99-457), services were extended to include infants and toddlers, resulting in areas of development that were not necessarily targeted by the traditional measures for older children (Calhoon, 1997), such as motor and language development. Four significant reasons for assessing preschool children have been identified in the literature: (1) to identify children in need of further evaluation (screening); (2) to provide diagnostic evaluation (diagnosis); (3) to facilitate program and intervention planning (treatment); and (4) to monitor children’s progress (Appl, 2000).

In 1990, President George H.W. Bush and State governors established the first National Educational Goal (“Goal 1”) through the National Educational Goals Panel stating that by the year 2000 all children would be able to start school with the skills
ready to learn (Shepard, Kagan, & Wurtz, 1998). While the goal’s intent was to help children, it was inherently problematic because the advisors, which included experts in early childhood development and assessment, to the National Educational Goal Panel could not locate data on or methods to measure young children’s development prior to school entry. As a result, in 1994, the United States Congress instructed the advisors to Goal 1 to “create clear guidelines regarding the nature, functions, and uses of early childhood assessments, including assessment formats that are appropriate for use in culturally and linguistically diverse communities, based on model elements of school readiness” (as cited by Shepard, Kagan, & Wurtz, 1998, p. 2). The Goals Panel was also instructed, “to advise and assist the Congress, the Secretary, the Goals Panel and others regarding how to improve the assessments of young children and how such assessments can improve services to children” (p. 2). Six general principles and recommendations were established to guide assessment practices for professionals and lawmakers. The principles state:

- Assessment should bring benefits for children; assessments should be tailored to a specific purpose and should be reliable, valid, and for that purpose; assessment policies should be designed recognizing that reliability and validity of assessments increases with children’s age; assessment should be age appropriate in both content and the methods of data collection; assessments should be linguistically appropriate, recognizing that to some extent all assessments are measures of language; and parents should be a valid source of assessment information as well as an audience for assessment results (As cited by Shepard, Kagan, & Wurtz, 1998, p. 5).
More specifically, the Panel suggested that early childhood assessment should cover the full range of development including cognitive, communication, motor, and social emotional and methods “of assessment should recognize that children need familiar contexts in order to be able to demonstrate their abilities” (p. 6). The Panel warned that the younger the child is, the more difficult it would be to obtain reliable and valid assessment data.

Despite the Panel’s principles and recommendations, there is ongoing need for the demonstration of valid, functional, and authentic tools to measure children’s development (Meisels, 1996). While parents, teachers, and professionals would like information about young children’s development, young children are by nature difficult to accurately assess. Assessing young children, from birth to five, is inherently challenging because it is the most rapid period of development (National Research Council and Institute of Medicine, 2000).

Tests administered at one time point may not yield a complete picture of the child’s development and needs. Young children do not necessarily have the experience or cognitive maturity to understand the goals of testing which can make the testing experience and interaction difficult or even impossible (Greenspan & Meisels, 1996). Young children also tend to demonstrate their knowledge better by showing or “doing” than by talking or writing (Shepard, Kagan, & Wurth, 1998). Moreover, development is impacted both by biological processes and powerful environmental forces making a stable prediction of developmental skills almost impossible (e.g., poverty, cultural, quality care giving). (See Luthar, 1999, National Research Council and Institute of
Medicine, 2000; Sameroff, 1994; and Sameroff & MacKenzie, 2003 for comprehensive reviews). The importance of these issues is clarified by examining the evolution of assessment practices. The following discussion will examine historical approaches to assessment and limitations to these approaches with young children and then conclude with a review on proposed practices for the assessment of young children.

*Historical Approaches to Assessment*

The first documented account of an “assessment” with children is from the 18th century (Sattler, 2001). Jean-Marc Gaspard Itard (1775-1838) was the first individual to investigate mental functioning in children as well as the differences between normal and abnormal cognitive and social functioning through his project, the “Wild Boy of Aveyron” (Hergenhahn, 2001; Sattler, 2001). Later, the French doctor, Esquirol, (1838) demonstrated degrees of cognitive impairment, which resulted in a classification system to diagnose mental retardation. He was also one of the first scientists to observe a relationship between language and cognition (Kelly & Surbeck, 1983). A codified discipline of assessment or testing however, really began seventy years later with Sir Francis Galton’s contributions to the discovery of the statistical concepts necessary to create the field of psychometrics (Sattler, 2001). As a result of these achievements, Galton has been termed the “father of mental testing” (Goodenough, 1949 as cited by Kelley & Surbeck, 1983). Meanwhile, Darwin’s studies of evolution (1859, 1872/1873) marked the first time a scientist demonstrated an awareness of the importance of infancy and early childhood on subsequent development. Darwin followed infants through maturation observing the sequence and rate of individual development.
Expansion of child assessment in the United States occurred several years later in 1899 when J.M. Cattell and Wilhelm Wundt founded the first psychological laboratory at the University of Pennsylvania. Cattell linked assessment to practical rather than purely philosophical implications by working from the premise that the assessment of mental ability could be studied experimentally (Sattler, 2001). During and following Cattell’s era, there were many more contributions to the measurement of mental and sensorimotor development in the United States, Germany, and France that culminated in the 1905 development of the Binet-Simon Scale, which evolved into the Stanford-Binet Intelligence Scale. This was the first intelligence scale that recognized variations in cognitive development across the age span and whose items acknowledged this principle (Sattler, 2001). Historical roots of child assessment are grounded in the notion that differential treatment of older and younger children is necessary in order to precisely determine whether a child performs below, at, or above the average level for his/her age (Bagnato & Neisworth, 1991).

During the early twentieth century, the United States experienced a rise in psychology clinics with a primary focus on child development, which led to an increased focus on assessment for young children (Kelley & Surbeck, 1983). Whereas testing for school aged children was well established by 1910 because of the Stanford Binet, tools for measuring preschool-aged children emerged several years later when Kulmann (1914); Burt (1921); and Yerkes & Foster (1923) created versions of intelligence tests for children under age three (as cited by Kelley & Surbeck, 1983). These tests did not have proper standardization procedures and did not demonstrate adequate reliability and validity (Kelley & Surbeck, 1983). Creating specific tasks based on age continued with
the Wechsler Intelligence Scale for Children in 1949 and can be observed in the revised versions of the Wechsler scale including the Wechsler Scale of Intelligence for Preschool Children (WPSSI-III; Wechsler, 2002).

A significant growth in testing for preschool children occurred after 1965 due to the increased role of the federal government in education. The 1964 Maternal, Child Health and Mental Retardation Act, the 1964 Educational Opportunity Act, and the 1965 Elementary and Secondary Education Act (Osborn, 1975 as cited by Kelley & Surbeck, 1983) funded programs such as Head Start and Follow Through to enhance educational and social opportunities for children from poor families. Additionally, congress enacted a provision that federally funded programs have a performance-based evaluation. As a result, new measures to assess young children greatly increased between 1965 and 1970 (Kelley & Surbeck, 1983). The tests were created to measure outcomes in four domains, social emotional or “affective” development, intellectual development, “psychomotor development” and particular subject achievement. One of the first tests developed was the Caldwell Preschool Inventory, which helped to inform preschool curriculum objectives. It is considered a precursor to the criterion referenced movement (Hoepfner et al., 1971 as cited by Kelley & Surbeck, 1983). Other notable examples include the Bayley Scales of Infant Development (Bayley, 1969; 1996; 2006) and the McCarthy Scales of Children’s Abilities (McCarthy, 1972).

Limitations to Historical Approaches

Professional standards including the National Association for the Education of young Children (NAEYC, 1991 as cited by Neisworth & Bagnato, 2000); the National Association of Early Childhood Specialists in State Departments of Education
(NAECS/SDE, 1991, as cited by Neisworth & Bagnato, 2000); the National Association of School Psychologists (NASP, Thomas & Grimes, 1995 as cited by Neisworth & Bagnato, 2000); the American Speech-Language-Hearing Association (ASHA, 1990, as cited by Neisworth & Bagnato, 2000); and the Association for Childhood Education International (ACEI, 1991, as cited by Neisworth & Bagnato, 2000) have voiced concerns with historical approaches to early childhood assessment. The concerns or limitations to traditional assessments are rooted in the professional literatures of early childhood development, early childhood assessment, early childhood education, and early childhood special education. The limitations can be characterized as practical, theoretical, and technical. Practical concerns are limitations that impede test administration in clinical settings. Theoretical concerns are limitations that occur as a result of an inherent relationship between child development theories and child assessment practices. Technical concerns are limitations that result from the psychometric properties of the assessment tools.

Practical Limitations to Traditional Assessment Procedures

Practical concerns are important to consider because each time an individual administers a tool for assessment she must quickly choose an appropriate, efficient, and easily accessible tool or test to use with the child. This is especially significant in light of data suggesting most states do not have a list of recommended assessment tools, making the choice of instrument the sole responsibility of the professional (Harrington & Tongier, 1993). More concerning, however, is that state departments of education do not feel as though there are a sufficient number of trained examiners to administer preschool
assessments (Harrington & Tongier, 1993). Practical limitations include professional views on the utility of traditional measures and the financial cost associated with test kits.

*Professional views of utility.*

Initial surveys of early childhood assessment practices revealed most professionals used a battery of diverse, but traditional standardized assessments, though they reported their batteries were inappropriate or inadequate for preschool and treatment recommendations (Kaplinski, Lidz, & Rosenfield, 1992; Schakel, 1987). Professionals have also indicated that traditional tools are less helpful for creating short-term objectives because of the small number of items at each age level (Bailey, Vandiviere, Dellinger, & Munn, 1987). Traditional measures also appear to pose challenges for children with disabilities. Teachers using the Battelle Developmental Inventory reported that less than half of the adaptation methods allowed by the testing manual enabled children to demonstrate their best performance (Bailey et al., 1987).

A later study (Bagnatto & Neisworth, 1994) (n=213 developmental school psychologists) assessed the acceptability of traditional preschool intelligence tests and their utility and value in determining early intervention eligibility. The survey respondents reported assessment data on 7,223 infants and preschoolers with developmental delays who were assessed for the first time to determine their eligibility for early intervention services. The authors found that 55% of the respondents (n=101) reported untestable data for over half of the children assessed which was defined as a failure to establish any of the following: basal or floor level; complete scores in all necessary subtests; clearly scoreable responses on a sufficient number of individual assessment tasks; interpretable standard scores without the necessity of “creatively”
altering the scoring procedures through the use of clinical judgment, or modifying the
administration procedures of tasks or stimulus properties of items; or without
accommodating the child’s unique response styles. Bagnato & Neisworth (1994) asked
professionals to report the reasons why traditional measures of intelligence and mental
development could not be completed in a valid way. The primary reasons for
determining that a child was untestable included, the child’s language deficits and the
language demands of the standardized measures; the child’s behavior that interfered with
performance on standardized test task; the rigidity of the standardized procedures that did
not allow modifications for the child’s impairments; and an insufficient number of tasks
to accommodate lower developmental competencies (high item floors).

Although a large percentage of children were determined to be untestable, over 90
percent of these children were declared eligible for early intervention services.
Psychologists reported using combinations of alternative strategies 80% of the time to
determine eligibility. The most frequently used alternative options included parent
interviews (58%), play-based assessment of interactions with toys and people (44%), and
parent-child observations in natural settings (30%). Psychologists also reported using
curriculum-based developmental scales and norm-based criterion referenced batteries
22% of the time (e.g., Brigance Diagnostic Inventory of Early Development). A
modified Q-sort procedure was used to code open-ended responses about the
professionals’ thoughts of norm-referenced conventional assessment measures. Only 4 %
of respondents thought norm-based, standardized intelligence and developmental tests
were appropriate for infants and preschool children with developmental delays. Eighty-
two percent of the respondents commented on the unrealistic nature and limitations of
standardized procedures with young children. Limitations included the ineffective “downward extension” of psychoeducational methods for young children, high item floors, lack of developmental sequencing, lack of utility and value for intervention programming, and the failure to include children with disabilities in the norm group. Respondents felt that assessment of young children demanded a multidimensional, team-based approach. Traditional measures were viewed as constraining, unattractive, unmotivating, too lengthy, and unresponsive to individual child differences. Professionals indicated the standardized administration of the assessment conflicted with the natural behavior and learning in child development. They viewed traditional assessment tools as misrepresenting information about young children’s functioning.

Non-western developmental psychologists have also voiced concern with the ability of standardized norm referenced tools to capture children’s developmental skills (Liao, Wang, Yao, & Lee, 2005). For example, items on standard tools may not address developmental skills related to specific cultures, such as the use of chopsticks for eating and feeding or particular cultural language development. The same concern, however, can be applied to the United States, where children from all cultures, and immigration statuses, require assessment. Moreover, standardized tools do not necessarily have adequate norms for the various sub-cultures within the United States (Fewell, 1991). A child from a recently immigrated Taiwanese family should not be necessarily compared to a child with Chinese decent, whose parents were born and raised in the United States.

_The expense of maintaining standardized toolkits._

Although limitations related to cost could be easily ignored because they are not inherent problems with the specific tools, they are nevertheless important to consider
because they impact the validity of the test results, which in turn impacts decisions made about and for young children. Formal assessments can be expensive (Shephard, Kagan, & Wurtz, 1998). Test kits contain multiple small items that are easily lost and costly to replace (Liao, Wang, Yao, & Lee, 2005). In lieu of replacing items with the proper test stimuli created by the test manufacturer, agencies with inadequate funding for assessment may be tempted to replace test items with apparently similar items. This small exchange could have larger consequences on the test administration if the exchanged item does not function equally across children impacting the standardization of the test. Formal testing also mandates a specific testing environment. It may be difficult for professionals working in systems that do not exclusively serve young children to find the appropriate space to accommodate these children during the formal test taking time. A child who is not provided a child size chair or desk will likely have a more difficult time demonstrating his typical skill set than a child who is seated adequately (McLean & Crais, 2004).

**Theoretical Limitations to Traditional Assessment Procedures**

The theoretical limitations to traditional assessment reflect the paradigm shift from an emphasis on linear theories of child development to an emphasis on dynamic and interconnected theories of child development (Linder, 2007).

*Historical linear view of development.*

Traditional or historical theories of development viewed development as a linear process, the child advanced through stages of development, building on prior skills and achievements (Erickson, 1963; Mahler, Pine, & Bergman, 1975; Piaget, 1962). Contemporary models of development, however, reveal the complexity of human growth
and suggest there are multiple levels of influence and multiple biological and environmental factors that interact to impact development (Bronfenbrenner, 1977, 1986; Meisels & Provence, 1989; Sameroff, Seifer, Barocas, Zax, & Greenspan, 1987). While development can be assessed separately through different measures, a child’s developmental process is no longer considered independent but interdependent (Greenspan & Meisel, 1996). For example, a child with a limited emotional vocabulary could become agitated when he/she realizes that he/she cannot successfully piece together a train track, resulting in a tantrum, screaming, and kicking. An assessment of language development might yield a finding suggesting the child has trouble word finding. An assessment of social development might yield a finding suggesting the child has problems regulating behavior. An assessment, however, that recognizes the interdependence of development will take into consideration how the child’s language development impacts the child’s frustration tolerance and progression towards tantrums (Linder, 1993).

*System change of development as interdependent.*

Assessment approaches should reflect current models of child development (Losardo & Notari-Syverson, 2001). Historical approaches to assessment, however, do not reflect this shift towards an emphasis on development as a dynamic and complex process (Meisels, 1996). For example, the underlying assumption of norm-referenced or standardized tests is that a child’s abilities are stable over time and are independent of other developmental domains or context (Losardo & Notari-Syverson, 2001). Traditional testing methods tend to artificially divide development into categories where as in a child’s daily life she is more likely to use her skills in combination rather than in isolation.
(Eisert & Lamorey, 1996). The interdependence of developmental domains makes it difficult to create reliable measures of specific developmental functions for early childhood (Meisels, 1994). As a result, experts have noted that traditional assessment methods tend to focus on narrow behaviors for which there are already existing scales, overlooking aspects of development that are more difficult to measure such as social emotional development (Greenspan, 1996; Greenspan & Meisel, 1996).

An assessment tool, which fails to measure the construct (i.e., development) that it sets out to measure is defined as one that does not demonstrate construct validity (Cronbach & Meehl, 1995) and produces error in the assessment resulting in an assessment that does not truly capture the construct it is intending to measure.

**Technical Limitations to Traditional Assessment Procedures**

The technical limitations to traditional assessment methods include an increased likelihood in creating measurement error (Bracken, 1987, 1988, 2007; Greenspan & Meisels, 1996; Mesiels, 1994; Meisels & Atkins-Barnett, 2000); inadequate psychometric properties (Bracken, 1987, 1988; Bradley-Johnson, 2001; Fuchs, Fuchs, Benowitz, & Barringer, 1987; Flanagan & Alfanzo, 1995; Meltzer & Reid, 1994); and failure to accomplish “treatment validity” (Neisworth & Bagnato, 2000, p. 34, Neisworth & Bagnato, 2004) which is the ability of an assessment tool to assess developmental competencies and weaknesses that are similar to treatment and program goals.

**Traditional assessments and measurement error.**

Measurement error, which can occur for a variety of reasons, compromises the validity of the assessment because it introduces biases or unwanted influences. For example, if a child is a dual language learner, a test with even a minimal amount of
language-based instructions may impede the child’s performance (Shephard, Kagan, & Wurtz, 1998). If the child does not do well, it is not known whether the problem is due to a misunderstanding of the instructions or a true area of weakness. Bracken (2007) has termed unwanted sources of variation as “construct-irrelevant variation.” There are four common sources of construct irrelevant influences on assessment results which compromise the validity of the evaluation that relate to the examinee, the examiner, the environment, and the actual instrument. A thorough understanding of the unwanted sources of variation helps to explain young children’s unpredictable performance on standardized evaluations (Bracken, 1987, 1988, 2007). These particular variables should be recognized and controlled during the standardized administration of assessments in order to facilitate the validity of the assessment results.

*Influences of the examinee.*

Young children’s age and level of development influence how they interact with the assessment process which can significantly impact the results of the assessment in a way that would not be expected in school-aged children. Meisels (1994) has identified four developmental limitations of young children that can create error in assessment results. First, due to their unsophisticated language development, young children have a restricted ability to understand the evaluator or assessment cues. Second, young children’s primitive visual and perceptual motor abilities can hinder their ability to correctly respond to the test task which may then be interpreted as “cognitive” error (Bordignon & Lam, 2004). Third, some questions on traditional measures require complicated verbal, executive, and information-processing skills—skills that are not fully developed in early childhood. Fourth, as a result of unrefined or even undeveloped
executive processing and information processing skills, young children may have difficulty understanding the demands of the task or what is being asked of them.

Additionally, while older children have mastered the task of separation, young children have not reached this level of sophistication (Bowlby, 1988). During a traditional assessment, young children are expected to perform or interact with novel tasks, in a novel room, and with a strange adult which could unintentionally cause distress, fear, distractibility, or anxiety to perform biasing the results of the assessment (Bracken, 2007; Greenspan & Meisels, 1996; Meisels & Atkins-Barnett, 2000). The level of motivation could also have significant impact on performance (Bracken, 2007, Shephard, Kagan, & Wurtz, 1998). Unlike older children, young children are not aware of the importance of the test. Motivation could wane as they become tired or the task becomes more difficult. As a result, the child may discontinue putting forth a valid effort, hindering assessment results.

_Influences of the examiner._

There are also several characteristics of the examiner that could compromise the validity of the assessment (Bracken, 2007). While the influence of examiner characteristics are not limited to traditional measures, the impact may be experienced to a greater extent because the child is already outside of his/her comfort zone during a testing environment (Neisworth & Bagnato, 2000). The examiner’s approachability to the child could either support the child’s participation in assessment or distance and/or alienate the child. Similarly, the examiner’s physical presence can either facilitate the assessment process or distract the child with unnecessary visual or auditory distractions such as loud jewelry or bright colored accessories (Bracken, 2007).
The examiner’s training can also influence the results (Bracken, 1988). For example, an evaluator who is not well versed in the rules of test administration may fumble through the manual as he/she attempts to engage the child. While an older child might sit patiently waiting for the task, a younger child might disengage and become distracted by surrounding objects due to his/her immature executive functioning skills, which could negatively impact test performance. Alternatively, if the examiner is consumed by the test administration, he/she may fail to establish rapport with the child, creating an uncomfortable and possibly stressful environment that discourages child participation.

*Environmental influences.*

Bracken (2007) has also identified characteristics of the environment that can either be effective in controlling unwanted construct influences, or add to the “noise” of measurement error. An effective environment is defined as one that is cheerful, conveys safety, capitalizes on the child’s curiosity, and stimulates the child’s participation. Aspects of the environment that can impede the validity of the assessment include the size and placement of the furniture. Furniture that is not child-size risks frustrating the child and adding unnecessary discomfort. An assessment room that is visually over stimulating with window and excessive wall pictures can negatively influence children’s ability to perform to their potential. There are clearly many possible combinations of unwanted influences that can compromise the validity of the standardized administration of young children’s assessment. In addition to the examinee, examiner, and environmental influences, there can be inherent psychometric problems within the actual tests designed for young children (Bracken, 1987; 1988; Flanagan & Alfonso, 1995).
Inherent psychometric problems with traditional measures.

Researchers have demonstrated problems with the reliability and validity of preschool measures (Meisels, 1987; Shepard & Graue, 1993). Early estimates indicated that while there were likely over 300 existing preschool measures, the percentage of measures with documented reliability and validity was likely less than 10% (Sheehan, 1988). Problems with the tools tended to occur for children between the ages of 2 to 3 years old due to the technical characteristics of the assessment tools as well as the inherent difficulty in reliably measuring children’s functioning at this lower age range because skills are often inconsistently demonstrated (Bracken, 1982 as cited by Bracken, 1987). Bracken (1987, 1988) identified 10 reasons why similar tests for preschool children might produce different results because of the psychometric problems related to standardized, norm-referenced tests by comparing 10 preschool instruments used for educational placement decisions and individual diagnostic assessment.¹ These reasons include “floor and ceiling effects, differences in item gradients, differences between test norm tables, poor or low reliability coefficients, variability in construct definition and evaluation, and, a non representative norm sample” (Bracken, 1987, p. 155).

Tests can have “floor effects.” The floor of a test is the lower range of standard scores that can be obtained when answering only a few or no items correctly. In order for

¹ The tools reviewed included the Battelle Developmental Inventory (Battelle); Bracken Basic Concept Scale (BBCS); Stanford Binet Intelligence Scale, Fourth Edition (SB-IV); Columbia Mental Maturity Scale (Columbia); The Kaufman Assessment Battery for Children (K-ABC); The McCarthy Scales of Children’s Abilities (McCarthey); Miller Assessment for Preschoolers (MAP); Peabody Picture Vocabulary Test-Revised (PPVT-R); The Token Test for Children (Token); and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI).
a test to distinguish children who are delayed or disabled from other children, the test needs to have enough easy items to assess the low-levels. The test floor for the entire assessment tool and for subtests is commonly the weakest characteristic of preschool instruments, especially below the age of four (Bracken, 1987; 1988; Flanagan & Alfonso, 1995). Weak floors can cause significant differences between two tests, which could have serious impact on diagnostic classification and intervention planning.

Tests can also demonstrate “ceiling effects.” A limited ceiling occurs when a test does not have enough items to distinguish a “very able” child, from an average or high average child in the assessed skill areas. Significant differences can occur when a test designed for younger children (e.g., Wechler Scale of Intelligence for Preschool Children) also overlaps in age range with a second tool designed for older children (e.g., Wechler Scale of Intelligence for Children). The test intended for the younger child would demonstrate a limited ceiling whereas the test for the older child would demonstrate a stronger ceiling.

Tests can differ in “item gradient”, or how steeply items are arranged within a test, which is related and can impact the content validity of the test. In order to obtain content validity, the test must do a good job sampling the content it asserts to assess, across the full range of difficulty levels. Steep item gradients can result in large differences in standard scores for small changes in raw scores. For example, the raw scores of 9, 10, 11, and 12 on the Cognitive Domain subtest on the Battelle Developmental Inventory (1984), would be converted into standard score equivalents (with a mean of 100 and a standard deviation of 15) with a range of 86 to 109, representing a change in more than one and one-half standard deviations for an increase
in three raw score points (Bracken, 1987). Although there is limited information about the impact of item gradients in the assessment literature, it is a critical test characteristic for the examiner to evaluate because it allows discrimination among various developmental ability levels (Flanagan & Alfonzo, 1995).

Some times the norm tables are not the same between measures. The norm tables provide information about average child development. They allow the evaluator to compare a particular child’s development with an average child (Sattler, 2001). If the age brackets for the norm tables, referred to as the “anniversary dates” i.e., when the child moves from one age norm table to another, are not the same, results from different tests can be incomparable.

Differences between test performances can occur due to poor or low reliability. Tests with low reliability have larger measurement errors resulting in wider confidence intervals where the true score lies. Additional reasons why two seemingly similar tests may produce dissimilar results stems from the way skills are assessed across tests as well as the way the construct may be defined. For example, two tests may assess the same global skill but differ significantly in the procedures for assessing the skill resulting in vastly different scores. Alternatively, the construct purportedly being assessed may be defined differently across measures resulting in markedly different scores (Bordignon & Lam, 2004).

Finally, differences between tests can occur when the norm sample is not representative of the population. Fuchs, Fuchs, Benowitz, & Barringer (1987) analyzed technical manuals of 27 aptitude and achievement tests to determine whether children with disabilities were included in the development of the test norms, items, reliability,
and validity indices. The authors concluded that most tests were invalid for use with young children with disabilities. This is problematic because while some children are referred to assess giftedness, most children are referred to assess a perceived problem, rather than strength, in development (Bracken, 1987).

In an investigation comparing the Stanford-Binet IV (SB-IV) and the Battelle Developmental Inventory’s (BDI) ability to classify high risk preschoolers with very low birth weight or intraventricular hemorrhage, the authors found that while the two measures were highly correlated, e.g., demonstration of concurrent validity, and correlated with future achievement outcomes, e.g., demonstration of predictive validity; the SB-IV and the BDI gave two completely different clinical outcomes or dispositions (Saylor, Boyce, Peagler, & Ashmore Callahan, 2000). Specifically, the SB-IV failed to detect 87% of the children who were identified as delayed by the BDI. Results demonstrated that these children, if they had been assessed using the SB-IV scores alone, would have been ineligible for early intervention services.

Recent reviews of test properties.

Similar to Bracken’s (1987) findings of the limitations to the technical properties of intelligence tests for young children, Flanagan & Alfonso’s (1995) review of more recently revised intelligence tests for preschoolers demonstrated that little had changed with regards to the technical adequacy of intelligence tests. This was again specifically true for children at the lower end of the preschool age range. The test floors and the subtest item gradients, continued to be problematic for children under age four. While Bracken’s review of tests found no single measure technically adequate across the entire preschool age continuum, Flanagan & Alfonso (1995) showed the Bayley Scales of
Infant and Toddler Development-II and the Woodcock Johnson Cognitive-Revised to be technically adequate across most criteria for three years-olds.

*Reviews of behavior rating scales.*

The discussed review studies (Bracken, 1988; Flanagan & Alfonso, 1995) were limited to tests of cognitive abilities (Nagle, 2007). A review of thirteen social emotional behavior ratings scales, by parents of young children, so called “third-party” instruments, demonstrated a collective strong picture of the measures, but no test, across all subtests, were determined to be technically adequate (Bracken, Keith, & Walker, 1998). The reviewers warned that use of third-party instruments requires an awareness of the psychometric weaknesses by domains, sub-domains, and age categories of the tool because the weaknesses can impact results and goal planning for intervention. The standardization sample apparently was the most problematic for the thirteen preschool behavior rating scales. Problems included restricted geographic regions and small sample size. Small sample sizes and geographically non-representative samples makes it difficult, if not impossible, to generalize the standardization characteristics to the general population.

*Reviews of instruments for children birth to three.*

The reviews above, however, only pertained to older preschool children, ages three to six and did not address test use with children under three years old. An investigation of cognitive development measures for use with children under three presented concerns about the technical adequacy of all the tests reviewed including the Battelle Developmental Inventory, Bayley Scales of Infant and Toddler Development II, Cognitive Abilities Scale, Leiter International Performance Scale-Revised, Mullen Scales
of Early Learning and Stanford-Binet-Fourth Edition (Bradely-Johnson, 2001). There were several frequently noted problems across measures. The norm samples consistently were not representative or even known across age levels because the norm characteristics were typically reported for the entire sample. Most of the norm samples did not include separate norms for children with developmental disabilities, making it impossible to assess atypically developing children.

A second common problem noted was the lack of stability data reported by age categories. Stability data was either not adequately demonstrated or was not reported. Test manuals also failed to demonstrate or report data on predictive validity, limiting the evaluation results to a particular time and context. Similar to cognitive measures designed for the older continuum of young children, the measures for the youngest children demonstrated inadequate test floors, making it difficult to diagnose mental retardation, resulting in the possibility of inflated scores.

Tests for very young children also demonstrated steep item gradients making it hard to detect or discriminate between performances below or above average. Many of the tests had drastic score changes for children at the cut-off age score within an age table. For example a child 17 months, 31 days old who obtained a raw score of 21 on the Cognitive Domain of the Bayley-II would receive a standard score of 91 (the average is 100 and the standard deviation is 15). If the child was tested on the next day at 18 months and obtained a raw score of 21, the child would receive a standard score of 65.

The limitations to traditional measures outlined above are theoretical, practical, and technical. To summarize, standardized measures have been criticized for failing to provide a holistic and in depth understanding of development, especially for children
with disabilities (Linder, 1990). Traditional testing artificially divides development into separate skills when in reality children use developmental skills in combination (Paget, 1983). Additionally, traditional measures can produce error or biases into the assessment process that can significantly impact test results (Bracken, 1987). For example, young children who may not have been exposed to formal schooling, may have difficulty performing to their ultimate abilities under the unfamiliar test taking environment so that an inappropriate or an incorrect response may just as likely reflect avoidant behavior or impulsivity as a skill deficit (Bordingnon & Lam, 2004).

Like many debates, the controversy surrounding traditional assessment has two camps (Bracken, 1994). One camp asserts that traditional assessments, particularly intelligence tests, can be used with preschool children as long as the examiner is aware of the test’s limitations, considers these limitations during test interpretation, and uses multiple sources of information, multiple instruments, across multiple contexts and constructs to substantiate or rule out hypotheses about the child’s intellectual functioning (Bracken, 1987; 1994; 2007; Bracken, Keith, & Walker, 1998; Flanagan and Alfonzo, 1995). The other camp asserts that traditional measures should never be used with preschool children, are inherently flawed, and should be “indicted, tried, and convicted for malpractice” (Neisworth & Bagnato, 1992, p.1). The limitations to traditional assessment do not suggest that professionals should discontinue evaluating young children. Routine, ongoing observation, and evaluation of young children benefit toddlers, preschoolers and their caregivers (Dichtelmiller & Ensler, 2004). Meisels (1996) argues that measurement activities should not be eliminated, but that “young children should be assessed carefully and in ways that are consistent with their
Experts in early childhood development and assessment emphasize assessment should take place in a natural environment, meaningful for the child (Puckett & Black, 2000). As children’s developmental needs are varied, the assessment process should evaluate all interdependent domains of development—cognitive, communication, motor and social-emotional (Neisworth & Bagnato, 2000).

**Best Practices for the Assessment of Young Children**

In the last ten years, organizations representing young children including the National Association for the Education of Young Children (NAEYC) and the Division of Early Childhood (DEC) have published recommended developmentally appropriate practices in early childhood assessment and intervention programs (See Bredekamp, 1997; Neisworth & Bagnato, 2000) based on the laws and regulations mandated by the Individuals with Disabilities and Education Act and the evidence supported in the literature. The recommendations contain two critical elements. First, parents and or family members should play a significant role in the assessment process. Second, the assessment methods and materials should accommodate the child’s unique development and/or disability (Neisworth & Bagnato, 2000). Above and beyond, assessments must be developmentally appropriate. The term developmentally appropriate has been operationalized to include eight qualities: “useful, acceptable, authentic, collaborative, convergent, equitable, sensitive, and congruent” (Bagnato & Neisworth, 1999 as cited by Neisworth & Bagnato, 2000 p. 19).

“Useful” refers to the ability of the assessment to demonstrate its purposes of evaluating, monitoring progress, and identifying children in need of services. A useful assessment is one that demonstrates “treatment validity” or an “essential similarity or
linkage among program goals, individual child objectives and the developmental competencies that are assessed” (Neisworth & Bagnato, 2000, p. 19). The “acceptability” of assessment indicates that the “methods, styles, and materials for assessment [are] mutually agreed upon by families and professionals” (p. 20). The assessment should be able to detect changes in behavior that are noticeable to caregivers as well as in early childcare environments. “Authenticity” refers to using natural environments and objects in children’s daily living to obtain a true appraisal of the child’s skills and needs. Assessment procedures should facilitate teamwork or “collaboration” between and among professionals and caregivers: “assessment must promote the concept of parent-professional decision making in which tests do not make decisions-people do” (p. 20). “Convergence” refers to collecting information about the child across professionals involved with the child, in multiple contexts, home and school. It is thought that the “pooling (or convergence) of several perspectives (family, professional) provides a better information base” (p. 20). “Equity” refers to the accommodation of individual differences in the test materials, procedures, and analysis of results. “Sensitivity” refers to the ability of the assessment to capture small increments of change so that progress can be demonstrated. Finally, “congruence” refers to the mandate that materials are designed for and “field-validated” (p. 21) on a wide range of children who will be assessed, from typical development, to children with mild and severe disabilities.

*Play Assessment Addresses the Limitations of Traditional Measures*

Currently there are two types of assessments available for use with young children. The first are the traditional, formal assessment tools described above. They provide standardized scores that can be helpful obtaining federal, state, and individual
private insurance funding (Linehan, Brady, & Hwang, 1991). Many states require the use of norm-referenced standardized results in their guidelines for funding. The second type of assessment is a play-based assessment. Standardized scores are not derived, however developmental levels are identified (Linder, 1990). Play assessments are often completed using a team-based approach. Team approach to assessment has been found to reduce miscommunication between developmental disciplines and increase consensus on specific child issues (Wolery & Dyke, 1984). Results can also be used for intervention and programming (Calhoon, 1997). Play assessments have likely emerged due to two factors. First, research in child development has provided evidence that play is a useful "yardstick" (Esret & Lamorey, 1996, p. 230) to understand the progression of development. Second, as described, professionals have voiced concern with the limitations to the traditional standardized methods for young children asserting that standardized tools are not functional, with techniques that are not appropriate for children with disabilities (Neisworth & Bagnato, 2000). In contrast, during a play assessment, "demands are not made on the child in terms of what, how, when, or how long they demonstrate skills" (Fewell, 1991, p. 169). Though experts have championed the use of play assessment with children (Fewell & Rich, 1987), they caution using play assessment procedures before empirical validation (Fewell, 1991).

The function of play.

It is often said that young children’s natural work is play; it is through play that children learn to explore themselves, others, and the world (Klein, Wirth, & Linas, 2003). Although it is a difficult action to succinctly define, early childhood experts agree that play is intrinsically motivating, enjoyable, spontaneous, and dynamic (Packer, Isendberg
In play, a child must call on all developmental skills to sustain play. Play has been labeled the “window into the child’s world” (Esert & Lamorey, 1996, p. 221) because it provides a demonstration of the interrelatedness of cognitive, affective, language, and motor development. Play has been shown to contribute to the acquisition of social, emotional, cognitive, linguistic, and motor skills (Moyles, 2005).

**Rational for play.**

Play has been studied across many disciplines, anthropology, art, evolution and biology (Ailwood, 2003). It is generally believed that play is correlated to the progress of all realms of development for both typical and atypically developing children (Casby & Ruder, 1983; Moyles, 2005). Play can occur independently or with other children and it has been found that when children are allowed to initiate play, particularly with other children, cognitive development is enhanced (Gmitrova & Gmitrov, 2004). Play allows children to self-prepare for the skills necessary to begin and sustain school. Frequent engagement in specific types of play is associated with an increase in readiness skills for kindergarten (Long, Bergeron, Leicht Doyle, & Gordon, 2005). Encouragement of play through the provision of toys in a primary care setting has even been found to reduce the amount of referrals to early intervention services (Tomopoulos, et al., 2006). Play has implications for the development of the child’s sense of self, and self in relation to others as it leads to the exploration of identity development including gender, cultural, and sexual identity (Broadhead & English, 2005; Hislam, 2005).

**Most Common Play Assessment Models Used By Professionals**

A play assessment is an evaluation that occurs in the context of play to determine the child’s current functioning and to determine if there are areas of concern in the child’s
development. Unlike traditional forms of early childhood assessment, there is a relatively short history and little information in the literature about the forms of play-based assessment. Additionally, the limitations of alternative, or play assessments have not necessarily been sufficiently examined (Barnett & Macmann, 1992, Bracken, 1994). Kelly-Vance & Oliver Ryalls (2007) assert there are only three forms of play assessment that have been described thoroughly enough to use in practice. They are the Play Assessment Scale (PAS) (Fewell & Rich, 1987); Play in Early Childhood Evaluation System (PIECES) (Kelly-Vance & Oliver Ryalls, 2005); and Transdisciplinary Play-Based Assessment (TPBA) (Linder, 1993, 2008).

*Play Assessment Scale.*

The Play Assessment Scale (PAS; Fewell & Rich, 1987) is intended for use with children from 2 to 36 months. It is a 45-item observational scale that is developmentally sequenced, and organized into eight age ranges with particular toy sets. The child is given a set of toys based on his/her age level and the administrator observes and scores the child in play. The toy sets are changed several times during the assessment to illustrate a wide range of play skills. The assessment begins with an observation of the child in “spontaneous,” free play, followed by an observation of the child during a facilitated play session, the child’s behaviors are then coded according to the scale and a play age is determined. The play age is established through observation of the play behaviors during the spontaneous play and is created by converting a raw play score to the child’s play age. There are not standard requirements for the administrator, such as a mandatory specific degree or training certificate, though the authors suggest the
administrator should have a comprehensive understanding of the PAS items and child development.

*Play in Early Childhood Evaluation System.*

The Play in Early Childhood Evaluation System (PIECES; Kelly-Vance & Oliver Ryalls, 2005) is the most recently developed approach. There does not appear to be a specified age range but presumably it is intended for children under the age of six or seven. It is a derivative of Linder’s Transdisciplinary Play Based Assessment (Linder, 1993) and focuses exclusively on cognitive development. PIECES involves observation of the child during free, independent play and it can be completed in any childhood setting. The PIECES coding guidelines consists of 13-items based on the empirical literature on the development of play. The information obtained during the assessment is then compared to the norms of typically developing children to determine areas in need of intervention.

*Transdisciplinary Play Based Assessment.*

Transdisciplinary Play Based Assessment-II (TPBA; Linder, 2008) is the revised version of the Transdisciplinary Play-Based Assessment (Linder, 1993) and is intended for the assessment of children from 6 to 72 months. It is the most widely used play assessment as well as the most internationally recognized (Kelly-Vance & Oliver Ryalls, 2007). The author’s intent is to assess the child’s developmental skills as well as the underlying developmental process, interaction patterns, and learning style (Linder, 1993, 2008). TPBA-2 is implemented by a team, which consists of the representative child development disciplines (e.g., speech and language, motor, social emotional, and educational) as well as the parent. The team observes the child during play with a play
facilitator, a caretaker, and a peer. TPBA-2 provides developmental guidelines to analyze the developmental level, learning style, interaction style, adaptive behaviors, and other relevant developmental behaviors. A developmental age as well as a score of functionality can be established for each of the developmental domains- cognitive, communication, motor, and social emotional. The TPBA-2 also has a hearing and vision screen that can be used across all covered ages. Like the PAS, there are not specific training requirements for the administration of the TPBA-2, however it is intended to be used by child development professionals and training is recommended.

Research on Play Assessment

Although there is ample theoretical and professional support for play as a vehicle for early childhood assessment, there is relatively few published research on play assessment (Eisert & Lamorey 1996). One exception is Farmer-Dougan & Kaszuba (1999) who examined the relationships between a general play-based assessment tool, the PLAY observation system, with the standardized assessment of cognitive and social skills through the Battelle Developmental Inventory (BDI) and the Social Skills Rating Scale (SSRC) in 3-5 year-olds. The authors found that play behaviors could predict scores on the BDI and on the SSRC and concluded that play-based assessments could provide a “meaningful” assessment of cognitive and social abilities. The authors did not investigate a relationship of play assessment and standardized tools with other developmental domains such as motor or language development.

Research supporting PAS.

There are three studies supporting the validity of the PAS. Fewell & Rich (1987) investigated the construct validity of PAS compared with six other standard
developmental assessments of language, cognition, and social skills. With one exception, the authors reported high correlations between the PAS and the six developmental assessments \((r = 0.80-.94)\). The authors hypothesized that the low correlation in the language domain \((r = 0.28)\) was the result of the sample of children who were non verbal. In a similar, but distinct study, Finn & Fewell (1994) expanded evidence for the construct validity of the PAS. They found that behaviors coded by the PAS were highly correlated with measures of receptive, expressive, and nonverbal language \((r = 0.87-.96)\). Finally, Esiert & Lamorey (1996) demonstrated a significant relationship at twelve and twenty months between the PAS and the Gesell Developmental Schedules \((r = 0.45-.56)\) and the Mullen Scales of Early Learning \((r = 0.82-.90)\). Stepwise regression was used to determine whether the high correlations between the PAS and the Mullen were attributable to a “third factor correlation,” chronological age. The authors found that the behaviors coded by the PAS were more predictive of the Mullen subtest scores than chronological age.

*Research supporting PIECES.*

Two published studies have reported on the reliability of the PIECES model (Kelley-Vance, Ryalls, & Gill-Glover, 2002; Kelley-Vance & Ryalls, 2005). As an extension of the 2002 study, the Kelley-Vance & Oliver Ryalls (2005) study of inter-observer reliability and test re-test reliability reported an inter-observer reliability of 90% for typically developing and 100% for exceptional children. The authors also found that the PIECES yielded a significant, but moderate, test-retest reliability for typically developing children \((r = 0.48, p = 0.015)\) and similar, though non-significant, test-retest reliability for exceptional children \((r = 0.575, p = 0.177)\).
Research supporting TPBA.

Although there is professional interest in Transdisciplinary Play Based Assessment (Eisert & Lamorey, 1996), there are few peer reviewed data on the reliability and validity of the original TPBA. The newest version of TPBA (TPBA-2) has no published studies to date. One published study of the original TPBA compared the social validity of TPBA to a multidisciplinary, standardized assessment. The authors defined social validity as “the validation of our work by consumers” (Myers, McBride, & Peterson, 1996 p. 103). The researchers found that compared to the standardized multidiscipline assessment, professionals involved in the TPBA process had a greater positive perception of the assessment, team meetings, and written reports. They also reported that there was a larger congruence between team members in transdisciplinary play-based assessment versus standardized assessment decisions. Professionals indicated that overall, the TPBA evaluation took significantly less time to complete and was more frequently completed within the time frame of IDEA requirements. Finally, written reports from the TPBA evaluation included more functional information than traditional reports that could easily be converted into useful recommendations.

A second published study completed by Kelly-Vance, Needelman, Troia, & Olvery Ryalls (1999) investigated the relationship between the Bayley Scales of Infant Development-II and the original TPBA on cognitive assessment and identification of children eligible for special education services. Child participants were primarily from middle-to upper middle-class Caucasian families. The authors created standardized age-equivalent scores using the developmental age tables of the TPBA (Linder, 1993) to compare with the Mental Developmental Index (MDI) of the Bayley-II. They found that
the play assessment yielded significantly higher scores than the MDI and offered three explanations. First, children may be able to demonstrate their optimum skill during a play assessment and may feel more restricted during a standardized test, specifically, “behavioral difficulties and noncompliance were more likely to negatively affect the BSID-II score” (p. 9) than the play assessment score. Second, the play assessment may be more subjective and as a result, influenced by the rater. Third, the authors questioned whether the play assessment had an “enabling aspect” due to the adult facilitation of play during the evaluation. The authors also looked at the percentage of children who would qualify for services. They found that 35 of the 38 children received the same eligibility decision on both assessment techniques while the remaining three children would have only qualified using the Bayley Scales of Infant Development-II. Finally, the authors reported a larger correlation between the Mental and Physical Developmental Indices of the Bayley-II (r = .87; p <.01) than between the Mental Developmental Index of the Bayley-II with the Cognitive Domain of the TPBA (r = .74) and the Physical Developmental Index of the Bayley-II with Cognitive Domain of the TPBA (r=.74).

It would be expected, however, that the two indices of the Bayley-II, while purportedly measuring different domains of development, would be highly correlated as the two indices share many items in common. While findings from Kelly-Vance & colleagues (1999) help to inform an understanding of the function of play in assessment, results cannot be generalized to the standard TPBA process as the comparison in this study relied solely on the cognitive domain of the TPBA and did not use the transdisciplinary nature of the assessment. Additionally, use of a team to assess a child, as communicated by the TPBA process, was not employed to generate cognitive scores.
The most recent and only other published study on the original TPBA was completed by Uys and colleagues in South Africa (Uys, Alant, & Lloyd, 2005). The researchers used the original TPBA to demonstrate the convergent validity of a daily behavior-rating index, the Daily Multiple Measurement Instrument (DMMI).

There are also several unpublished doctoral dissertations and one unpublished manuscript concerning the original TPBA (Cornett, 1998; deBruin, 2005; Friedli, 1994; Linder, Goldberg, & Goldberg, 2007). Linder, Goldberg, & Goldberg, (2007) provided evidence for the interconnections and relationships between the domains of development. The authors surveyed experts corresponding to the developmental disciplines and asked them to rate the impact of their discipline on other domains of development as well as the impact of the other domains of discipline on their own expert domain area. Algorithms and graphic representations from the professional responses substantiated the transdisciplinary nature of development.

Friedli (1994) demonstrated concurrent validity of the original TPBA with the Battelle Developmental Inventory. She reported a strong association in the form of Phi coefficients (Phi coefficient =.90) in the ability of the TPBA to determine a child’s early intervention eligibility as well as in its ability to create a similar strengths and weakness child profile (Phi coefficient = .76). The author also demonstrated the test-retest reliability of the original TPBA and content validity of the four domains of development measured by TPBA, speech and language, social emotional, cognitive, and motor development.

Cornett (1998) corroborated Friedli’s demonstration of the content validity of the original TPBA. The author found that the levels of agreement between professional
judges ranged between 70% and 85% when estimating a child’s chronological age range. Despite the consistency between the judge’s estimate of chronological age, the author found that the professional judges produced a variable pattern of consistency when identifying specific play behaviors, suggesting that while professionals do not necessarily use the same behaviors to make judgments about the child’s level of functioning within the four domains, they are able to consistently come to similar conclusions using TPBA about the child’s functioning in order to make decisions about intervention planning.

deBruin (2005) investigated the association between the original TPBA and the Battelle Developmental Inventory-II on categories of eligibility and non-eligibility and found a non-significant association between TPBA and the BDI-II. Results suggested, however, that through the use of a social validity questionnaire, the professionals involved in the assessment believed TPBA to be more appropriate at identifying children with developmental delays, particularly speech delays. The lack of agreement between the TPBA and the Battelle Developmental Inventory could have stemmed from the composition of the team administering the play assessment. The TPBA team in this study consisted of a preschool teacher, a paraprofessional, a speech and language pathologist, and a parent, whereas the individuals completing the Battelle assessments were two Licensed Specialists in School Psychology. More research is necessary to determine the viability of play assessment compared to traditional standardized measures.

There must be a universal language to describe and communicate the effectiveness of a measurement tool that informs assessment. The field of early childhood assessment presents many challenges. While numerous standardized tools exist to separately measure each domain of development, these tools pose psychometric
and theoretical problems. Yet, alternative measures, such as play assessment, do not have a history of evidence supporting it. Experts concede that “the challenge in contemporary assessment is to know what to measure, how to measure it, and whether the measurements are meaningful” (Wasserman & Bracken, 2003, p. 43). Two ways of characterizing whether an assessment tool is meaningful is through the psychometric characteristics of validity and reliability. Validity refers to the extent to which the tool measures what it intends to measure (Bordons & Abbott, 1999), the meaning of test scores (Cronbach & Meehl, 1955), or the “appropriateness with which inferences can be made on the basis of tests results” (Sattler, 1988 as cited by Hammill, Brown, & Bryant, 1993, p.11). Reliability is the degree to which the test consistently produces similar scores.

**Validity**

Validity is a general term with historically several specified types. Cronbach & Meehl (1995) described four types of validation: predictive, concurrent, content, and construct validity. The first two have been considered together as criterion-related validity. The most recent *Standards for Educational and Psychological Tests* developed jointly by the American Educational Research Association (AERA) and American Psychological Association (APA) (1999) continues to suggest that test developers investigate the three forms of validity, criterion-related, content, and construct. By extension, alternative measures of assessment should be able to demonstrate validity in order to provide administrators and consumers the confidence that the test is measuring what it purports to measure (Hammill, Brown, & Bryant, 1993).
Criterion-Related Validity

Criterion-related validity is the degree of relationship between test scores and a designated criterion, usually another test that presumably measures the same construct (Hammill, Brown, & Bryant, 2003). There are two types of criterion-related validity, predictive and concurrent. The predicative validity of an instrument illustrates the test’s ability to estimate or predict a future behavior outcome, such as the need for special education services at a later age. Predicative validity allows one to determine, for example, that a preschool child, who scored low on a developmental test or school readiness measure, eventually required special services. Concurrent validity demonstrates equity between test measures. It suggests that performance on one test is comparable to performance on another test of the same construct (Hammill, Brown, & Bryant, 1993).

Content Validity

Content validity is the degree to which a particular domain, or element of a test, is sampled adequately (Bracken and Wasserman, 2003; Hammill, Brown, & Bryant, 1993). Content validity is typically established through the use of an expert panel of judges who review the test content. It has been noted that the differences between tests that purport to measure the same construct can be accounted for by the differences in the test content (Bracken & Wasserman, 2003).

Construct Validity

Construct validity is used to estimate the extent to which a test is able to measure a variable that is not directly observable such as intelligence or self-esteem. Test developers create tools to measure behavior that is believed to represent the “construct”
(Bordons & Abbot, 1999); if the construct is defined too narrowly or broadly, construct validity can be threatened. A construct that is too broadly defined can result in the construct irrelevance phenomena that Bracken (1987, 1988) warns against. To review, construct irrelevance refers to the extent to which test scores are impacted by unintended factors.

Current Forms of Validity

To the extent possible, evaluators should choose measures with acceptable levels of reliability and validity (Bagnato & Neisworth, 1991). However, as discussed above, many preschool assessments measures do not meet adequate technical standards, which can threaten all forms of validity (Bracken, 1987; 1988). This presents additional problems for professionals when attempting to evaluate the validity of a measure. Neisworth and Bagnato (1992) identified five myths related to the use of intelligence tests with children under the age of three which demonstrate the complexity and challenges presented to the evaluator when sifting through the technical components of a measure. The first myth relates to the principle of construct validity and states, “professionals know what early intelligence is and agree it can be measured” (Neisworth & Bagnato, 1992, p.2). In other words, professionals assume that intelligence is a uniformly defined construct, yet intelligence is defined in different terms at different ages for children under three (Meisels & Atkins-Burnett, 1999). Moreover, for young children, “intelligence” is not an independent entity that can be realistically measured as it is connected and impacted by the other domains of development, emotional, language, and motor (Linder, 1993; 2007; 2008; Meisels & Atkins Burnett, 1999; Meisels & Atkins-Burnett, 2000).
Attempting to measure the validity of the construct of intelligence in young children becomes a circular pursuit: it makes the assumption “that we know what the construct is that we are measuring before we have even been able to define the construct” (Meisels & Atkins-Burnett, 1999, p. 24). The second and third myths are directly related to an assumption of validity and reliability and states, “research supports the measurability, reliability, and diagnostic utility of early intelligence testing” (Neisworth & Bagnato, 1992, p. 3) and that “early intelligence tests have predictive validity” (p. 5). The problem lies again in the nature of child development. Intelligence in young children represents more of an accumulation of current skills and behaviors in particular contexts than a predictive indicator for later abilities. Additionally, the assumption underlying predictive validity is that the construct is uniformly defined and stable over time. However, it is well documented that environmental factors such as the home environment, community resources, access to health care, and life events, have more “predictive power” on future development than early intelligence tests (Sameroff, Seifer, Baldwin, & Baldwin, 1993; Sameroff, Seifer, Barocas, Zax, & Greenspan, 1987). The fourth myth concerns the assumption of content validity, particularly for children with special needs, who are most typically the children, referred for evaluation. It states, “standardized administration procedures provide a reasonable and representative assessment of young children with special needs” (Neisworth & Bagnato, 1992, p. 8). Unfortunately, as Bracken (1987, 1988) outlines, test procedures and content do not necessarily account for a child’s specific disability. The tests do not necessarily sample a broad enough content of skills to adequately demonstrate the child’s abilities. Neisworth & Bagnato (1992) challenged that these instruments “really measure the child’s disability
rather than ability” (p. 8). The fifth myth is related to the general assumption of validity, or the ability to make accurate inferences from test scores and states, “psychologists and other practitioners value and use intelligence tests to identify young children in need of early intervention” (Neisworth & Bagnato, 1992, p.9). The evidence against this myth is revealed through a survey conducted by Bagnato & Neisworth (1994) where 43% of psychologists and other practitioners specializing in young children found early intelligence tests to be meaningless, i.e., without value, for the purposes of early intervention planning.

Taken together, the limitations of preschool measurements as well as the inherent problems associated with evaluating an assessment tool’s value through conventional standards of validity suggests there needs to be additional means of evaluating the accuracy of assessment tools designed for use with young children. Indeed, Meisels & Atkins-Burnett (1999) argue that we “need to rethink our methodology for learning about children’s growth and development” (p. 25). As a result, the dimension of validity has expanded to include newer forms that deserve equal attention in the literature, though limited evidence exists documenting their support.

Current Forms of Validity

Newer forms of validity include treatment validity (this has also been labeled treatment utility by Hayes, Nelson, & Jarrett, 1987), decision validity (Barnett & Macmann, 1992), and ecological validity (this has been termed social validity by Bagnato & Neisworth 1991; Wolery & Dyke, 1984).
Treatment validity.

Treatment validity or utility refers to the extent to which the assessment directly guides plans for interventions, or “the degree to which assessment is shown to contribute to beneficial treatment outcome” (Hayes, Nelson, & Jarrett, 1987, p. 963). Treatment validity is considered a demonstration of the pragmatic advantage of the tool. The question treatment validity or utility seeks to answer is - 'is this tool useful for intervention?' (Hayes, Nelson, & Jarrett, 1987; Nelson-Gray, 2003). It is considered a method of examining the quality of the assessment. A tool is considered to demonstrate treatment utility if it can show that the “treatment outcome is positively influenced” by the tool (Hayes, Nelson, & Jarrett, 1987 p. 964). Proponents emphasizing treatment validity or utility argue that the psychometric properties evaluate the structural aspects of a tool and do not measure the tool’s functionality.

Decision validity.

Decision validity is very similar to the concept of treatment validity. It takes on a slightly broader definition to include the diagnosis and treatment decision processes. Decision validity evaluates the appropriateness of using assessment data to make diagnosis or treatment decisions. The focus of decision validity is on “the interpretations and inferences that the user draws from the test scores, and the decisions and actions that flow from those inferences” (Angog, 1988 as cited by Barnett & Macmann, 1992, p. 432). Social validity is also similar to the notions of treatment validity and decision validity and has been used in the context specifically related to early childhood assessment (Wolery & Dyke, 1984). It is seemingly unknown outside of the early
childhood and early childhood special education literature. A more common and perhaps identifiable term is “ecological validity.”

*Ecological validity.*

Ecological validity was initially exclusively used to evaluate research methodology, e.g., the external validity, or the ability to make generalizations from the research protocol to real life contexts. The concept of ecological validity has been debated because there has not been a clear consensus on its definition (Schmuckler, 2006). Bronfenbrenner produced a “classic” definition stating “ecological validity refers to the extent to which the environment experienced by the subjects in a scientific investigation has the properties it is supposed or assumed to have by the experimenter” (Bronfenbrenner, 1977, as cited by Schmuckler, 2006, p. 421). Ecological validity involves maintaining a connection with real-life scenarios within the experimental context.

The field of neuropsychology has expanded the concept of ecological validity to evaluate the assessment context (See Chaytor & Schmitter-Edgecombe, 2003 for comprehensive review). This is a logical extension if one considers assessment as a micro experiment with one subject or participant. Researchers in neuropsychological assessment have defined ecological validity as “the degree to which test performance corresponds to real world performance” (Chaytor & Schmitter-Edgecombe, 2003, p. 185).

*Measuring ecological validity.*

There are two identified approaches to evaluate ecological validity: versimilitude and veridicality (Chaytor & Achmitter-Edgecomb, 2003). These approaches evaluate
how well the test captures everyday, functional capacities, rather than achievement of developmental milestones. In the context of early childhood assessment, the versimilitude approach would involve comparing a traditional developmental measure with a second tool that has been identified as simulating real world behaviors, such as adaptive behaviors. Hypothetically, according to this approach, an evaluator may be able to compare the WPPSI with the results of the Vineland Adaptive Behavior Scale.

Veridicality is another, though similar, approach used to evaluate ecological validity. It refers to the degree to which “existing tests are empirically related to measures of everyday functioning” (Franzen & Wilhelm, 1996 as cited by Chaytor & Schmitter-Edgecombe, 2003 p. 183). This involves relating traditional measures of developmental skills with measure of “real-world” functioning such as a clinician rating scale, chart review, or interviews with family members. For a young child, this could translate to a comparison of the traditional measure of development with parent ratings of the child, as the parent is the individual who knows the child the most and can estimate his/her child’s daily functioning (Glascoe, 2000; Glascoe & Dwarken, 1995; Young, Davis, Schoen, & Parker, 1998).

The newer forms of validity reflect a value on the functional capacity of the assessment process to contribute a more realistic picture of the child’s strengths and weaknesses through a developmentally appropriate evaluation resulting in favorable treatment planning or outcomes for parents, professionals, and schools rather than an emphasis on achievement of developmental milestones (Meisels & Atkins-Burnett, 2000). These new forms of validity evaluate the assessment’s ability to achieve a major and significant purpose of assessment in early childhood, which is to attain “reliable,
valid, and communicable descriptors of a child’s development that will provide a basis for describing and documenting appropriate treatment” (Simeonson, Huntington, & Parse, 1980 as cited by Fleisher, Belgredean, Bagnato, & Ogonosky, 1990, p. 13).

Critics of traditional standardized measures and conventional forms of validity warn that there is little to be gained through the analysis of test and subtest scales without placing them within a social context (Barnett & Macmann, 1992). Ideally, assessment tools for young children would demonstrate both conventional forms of validity as well as these newer or alternative forms (Hayes, Nelson, & Jarrett, 1987 p.972). While the newer forms or dimensions of validity have historical roots in psychometric theory, they have been neglected in the literature (Barnett & Macmann, 1992). With the exception of the neuropsychological assessment literature, there is little information regarding acceptable methods for evaluating these forms of validity, particularly for early childhood assessment. This latter problem echoes an historical challenge relating to a general lack of consensus on acceptable forms of research methodology to evaluate conventional forms of validity (Hammill, Brown, & Bryant, 1992).

**Purpose of the Current Study**

The purpose of the present study was fourfold. The first purpose was to examine the concurrent validity of the Transdisciplinary Play Based Assessment (TPBA-2) with two widely used norm-referenced measures. For children 1-36 months, the Bayley Scales of Infant and Toddler Development-Third Edition (Bayley-III; Bayley, 2006) was used; for children 36-68 months the Mullen Scales of Early Learning (MSEL; Mullen, 1995) was used. The second purpose was to examine the ecological validity of the TPBA-2. The third purpose was to determine whether the norm-referenced measure or play-based
measure resulted in facilitating the child’s most typical skills and abilities. The final purpose was to determine if a relationship exists between test score and test behavior and whether there was a difference by testing conditions.

There are no published studies that have investigated the concurrent validity of the TPBA-2, the Bayley-III, and the MSEL using all developmental domains and using the team approached outlined by Linder, 1993, 2008. There are no published studies that have investigated the ecological validity of the TPBA-2. There are no published studies that have investigated which assessment excels at eliciting the child’s most typical set of skills. Finally, while there have been investigations demonstrating that noncompliant test-taking behavior can significantly impact test scores in older children (Glutting, Youngstrom, Oakland, & Watkins, 1996, as cited by Sattler, 2001) as well as hypothetical assumptions that test taking behavior in young children can impact test scores (Bracken, 1994; Kelly-Vance, Needelham, Troia, & Oliver Ryals, 1999); there are no empirical studies investigating the relationship between young children’s test taking behavior and test score.

Research Questions

Concurrent validity.

Is there a statistically significant association between the Transdisciplinary Play-Based Assessment-2 (TPBA-2) and the Bayley Scales of Infant and Toddler Development-Third Edition (Bayley-III) on the developmental domains of cognitive, language, motor, social-emotional, and total development? See Table 1 for research goals and outlined statistical analysis for all research questions.
Is there a statistically significant association between the TPBA-2 and the Mullen Scales of Early Learning (MSEL) on the developmental domains of cognitive, language, motor, and total development?

*Ecological validity.*

Does the TPBA-2 demonstrate ecological validity as measured by the association between TPBA-2 and the Vineland Adaptive Behavior Scale-Second Edition (Vineland, II) on the domains of cognitive, communication, motor, social emotional, and overall development?

*Facilitation of typical skills.*

Is there a statistically significant difference between the ability of the TPBA-2 and the standardized developmental assessment to demonstrate the child’s most typical skills and abilities during the assessment as measured by parent evaluation of the child’s typical home behaviors on the Parent Rating Scale of Typical Behavior?

*Test-taking behavior.*

Is there a significant correlation between test score and a child’s test taking behavior as measured by the Test Taking Behavior Rating tool?

Is there a significant difference between the Test Taking Behavior Rating Scale Score during the administration of the TPBA-2, Bayley-III, or MSEL as measured by the Test Taking Behavior Rating Tool?
Table 1

*Research goal, statistical analysis, alpha, power, and effect size*

<table>
<thead>
<tr>
<th>Research goal</th>
<th>Statistical analysis</th>
<th>Alpha</th>
<th>Power</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concurrent Validity</td>
<td>Pearson Correlation</td>
<td>&lt;.05</td>
<td>.80</td>
<td>.30</td>
</tr>
<tr>
<td>Ecological Validity</td>
<td>Pearson Correlation</td>
<td>&lt;.05</td>
<td>.80</td>
<td>.30</td>
</tr>
<tr>
<td>Facilitation of Typical Skills</td>
<td>Dependent Sample T-test</td>
<td>&lt;.05</td>
<td>.80</td>
<td>.50</td>
</tr>
<tr>
<td>Test Taking Behavior and Test Score</td>
<td>Pearson Correlation</td>
<td>&lt;.05</td>
<td>.80</td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td>Dependent Sample T-test</td>
<td>&lt;.05</td>
<td>.80</td>
<td>.50</td>
</tr>
</tbody>
</table>

Note. Power and effect size are based on the standards outlined by Cohen, 1992.

*a* Based on sample size of 90 participants.

*b* Based on sample size of 65 participants.
CHAPTER TWO

Method

Participants

Forty children with and without disabilities (21 girls and 19 boys, mean age = 38.22 months, SD = 18.52 months, age range = 10-73 months, mode = 18, 51, 67 months) and 40 mothers were recruited through community advertising, newspaper advertising, “word of mouth,” and Internet group postings in a large western city. The sample included 15 (37.5%) children with no formal diagnosis or identified developmental delay, 7 (17.5%) children with Autism Spectrum Disorder, 4 (10%) children with problems with sensory integration, 2 (5%) children with Down Syndrome, 2 (5%) children with a speech and language delay, 1 (2.5%) child with a motor delay, 6 (15%) children with an “other” diagnosis or documented delay, and 3 (7.5%) children for whom a diagnosis or delay was not reported. “Other” diagnosis included visual impairment, 23Q deletion syndrome, Albinism, and dual diagnoses. In regards to race and ethnicity, the child sample included 30 (75%) Caucasian, 3 (7.5%) Asian, 0 (0%) African American, 1 (2.5%) Hispanic, 4 (10%) bi-or-multiracial children, and 2 (5%) children for whom ethnicity was not reported. The mean income for parents was $94,708 (SD = $4.37). In addition to family participation, 46 professionals in early childhood development (43 females and 3 males, mean age = 38.8 years, SD = 12.61 years, age range = 24-63 years) volunteered to participate. Professionals were recruited from the annual Summer Transdisciplinary Play Based Assessment Training Institute. The professional backgrounds included 12 (26.1%)
Early Childhood Educators, 8 (17.4%) Occupational or Physical therapists, 7 (15.2%) Graduate students, 6 (13%) Educational or school psychologists, 5 (10.9%) Speech and language pathologists, 3 (6.5%) Social workers, 3 (6.5%) “Other,” and 2 (4.3%) Clinical psychologists. Professionals’ years of experience in their discipline ranged from 1-38 years (mean = 12.88 years, SD = 10.93 years). Professionals’ years of experience using the Transdisciplinary Play Based Assessment ranged from 0-18 years (mean = 1.77 years, SD = 3.74 years). Three graduate level research assistants administered the standardized assessment tools. Each had completed five assessment courses; two general assessment courses, two focused on early childhood, and one focused on social emotional assessment. Additionally, each had participated for a full year at the University Assessment clinic including a proportion of time completed in the Early Childhood Clinic.

A power analysis was completed using the medium effect size and power standards of Cohen, (1992) and the G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007) which indicated that 90 participants were needed for associational data analysis and 65 participants were needed for differences data analysis to prevent Type I error.

All participants were English speaking. Caretakers signed consent forms to permit participation of her child. Professionals completing the Transdisciplinary Play Based Assessment Summer Institute signed consent for participation. Participants were treated in accordance with the “Ethical Principles of Psychologists and Code of Conduct” (American Psychological Association, 2002).
Materials for Measurement

Two well recognized standardized assessment tools as well as the Transdisciplinary, Play-Based Assessment-2 (TPBA-2; Linder, 2008) were selected. The Bayley Scales of Infant and Toddler Development- Third Edition (Bayley-III; Bayley, 2006) was chosen as the measure for children ages 1-36 months to compare with the TPBA-2 because it is the most widely used measure to assess young children in both research and clinical settings (Johnson & Marlow, 2006). It is also considered a comprehensive evaluation that sets “the standard for early childhood assessment” (Albers & Grieve, 2007). The Mullen Scales of Early Learning (MSEL) was used for children 37 months-69 months because the Bayley-III does not cover the full early childhood age range of birth to six. Although the MSEL can be used with younger children and demonstrates acceptable reliability and validity for children under the age of two (Bradely-Johnson & Johnson, 2007) it is not as commonly used with this age range in clinical settings. Table 2 presents the number of children in each age group and their corresponding average age and standard deviations. Table 2a presents the demographic information of the children by standardized assessment tool.
Table 2

*Child average age by age-group*

<table>
<thead>
<tr>
<th>Age group</th>
<th>Mean age</th>
<th>Standard deviation</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-36 months</td>
<td>23.05</td>
<td>10.36</td>
<td>19</td>
</tr>
<tr>
<td>37-72 months</td>
<td>51.95</td>
<td>12.49</td>
<td>21</td>
</tr>
</tbody>
</table>
Table 2a  

*Child demographics by assessment tool*  

<table>
<thead>
<tr>
<th>Assessment Tool</th>
<th>Bayley-III</th>
<th>MSEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Col%</td>
<td>N</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>13</td>
<td>68.4</td>
</tr>
<tr>
<td>M</td>
<td>6</td>
<td>31.6</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Asian</td>
<td>1</td>
<td>5.3</td>
</tr>
<tr>
<td>Caucasian</td>
<td>14</td>
<td>73.7</td>
</tr>
<tr>
<td>Multi-racial</td>
<td>2</td>
<td>10.5</td>
</tr>
<tr>
<td>Not reported</td>
<td>1</td>
<td>5.3</td>
</tr>
</tbody>
</table>

*a* Dual Diagnosis
<table>
<thead>
<tr>
<th>Disability</th>
<th>Bayley-III</th>
<th>MSEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autism</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Down Syndrome</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Gross Motor</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Sensory</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Speech/Language</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Other b</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>No Diagnosis</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>Not Reported</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

* b Dual Diagnosis, Global Developmental Delay, Potlocki Lupski Syndrome, 22Q3 Deletion Syndrome
The Vineland Scales of Adaptive Behavior, Second Edition (Vineland-II; Sparrow, Cicchetti, & Balla, 2005) was used as the target comparison of functional skills to determine whether the TPBA-2 demonstrates ecological validity. The Vineland-II is considered the “gold standard” measurement of adaptive behavior. The authors define adaptive behavior as the “performance of daily activities required for personal and social sufficiency” (Sparrow, Cicchetti, & Balla, 2005, p. 6). Adaptive behavior is considered to be age related and defined by the expectations or standards of other people. Those who interact closely with the child judge the adequacy of the child’s adaptive behavior. Adaptive behavior is also defined by usual performance rather than ability. A child may have the ability to demonstrate a particular skill but seldom use the skill when it is needed, the child’s adaptive behavior would be considered to be inadequate in this area (Sparrow, Cicchetti, & Balla, 2005).

One parent measure, the Parent Rating Scale of Typical Behavior, was used to assess whether the child’s performance was typical during the assessment. This measure was created specifically for use in the current study as no measure exists that quantifies a parent’s perception of his/her child’s typical performance elicited by a developmental evaluation. One behavior rating tool was used to assess the child’s behavior during testing, the Test-Taking Behavior Rating Scale. This measure was also created for the current evaluation.


The Bayley Scales of Infant and Toddler Development-Third Edition (Bayley-III) is an individualized, standardized assessment tool designed for children ages 1-42.
months. The norm sample is reported to have included 10% of children with developmental disabilities. It assesses five domains of development- Cognitive, Language, Motor, Social-Emotional, and Adaptive Behavior (Albers & Grieve, 2007). The Adaptive Behavior scale was not used in the current study. The language scale consists of receptive and expressive language subtests. The motor scale consists of gross and fine motor subtests. The social-emotional scale is completed by the caregiver. This scale also contains a sensory processing score. The five domains can be administered in any order, but within each subtest, the test administrator is instructed to follow item administration order (Bayley, 2006). A standard score and age equivalent score is derived for the cognitive, language, and motor scales. The standard score has an average of 100 and standard deviation of 15.

The Bayley-III Scale composite average internal consistency reliability coefficients range from $r = .91$ (Cognitive) to $r = .93$ (Language). Subtest average internal consistency reliability coefficients range from $r = .86$ (Fine Motor subtest) to $r = .91$ (Expressive Communication and Gross Motor subtests). Internal consistency coefficients for the Social Emotional Scale range from $r = .83$ to .94 for social emotional-emotional items and $r = .76-.91$ for the sensory processing items. Total test-retest stability correlation coefficients range from $r = .67$ (Fine Motor subtest) to .80 (Expressive Communication subtest).

The validity of the Bayley-III has been substantiated utilizing the Wechsler Preschool and Primary Scale of Intelligence-Third Edition (Wechsler, 2002, as cited by Bayley, 2006) $r = .71-.83$; the Preschool Language Scale-Fourth Edition (Zimmerman, Steiner, & Pond, 2002, as cited by Bayley, 2006) $r = .51-71$; the Peabody Developmental

**Mullen Scales of Early Learning.**

The Mullen Scales of Early Learning (MSEL) is an individualized, standardized assessment tool designed for children ages 1-68 months. The MSEL norm sample excluded children with disabilities. The MSEL has five scales: Gross Motor, Visual Reception (Cognitive scale), Fine Motor, Receptive Language, and Expressive Language. General Intelligence is reported to be assessed by the “Early Learning Composite,” consisting of all the scales but the Gross Motor Scale. The Gross Motor Scale is not used with children over three years-old. Results for each scale can be described by T scores ($M = 50, SD = 10$) and age equivalents. The Early Learning Composite (ELC) is described as a standard score ($M = 100, SD = 15$). Scale scores can be transformed to standard scores with an average of 100 and a standard deviation of 15.

The MSEL reports median internal consistency coefficients as $r = .92$ (Gross Motor); $r = .79$ (Visual Reception); $r = .75$ (Fine Motor); $r = .80$ (Receptive Language); $r = .80$ (Expressive Language); and $r = .91$ (Early Learning Composite). Test-retest reliabilities for the ages of 25 months to 56 months are reported as $r = .75$ (Visual Reception); $r = .79$ (Fine Motor); $r = .77$ (Receptive Language); and $r = .71$ (Expressive Language).

The MSEL has been validated utilizing the Bayley Scales of Infant Development-II. The four MSEL scales correlated with the Bayley Mental Development Index with a
range of r = .53 -.59 and with the Bayley Psychomotor Development Index with a range of r = .21-.52. The MSEL has also been validated utilizing the Preschool Language Assessment (Zimmerman, Steiner, & Evatt, & Pond, 1979), reporting a range of r = .72-.85 and the Peabody Fine Motor Scale (Folio & Fewell, 1983), reporting a range of r = .65-82.

Transdisciplinary Play Based Assessment-Second Edition.

The Transdisciplinary Play Based Assessment-Second Edition (TPBA-2) involves assessing the child in an informal play setting that contains manipulatives, representational toys, tactile and art materials, construction play objects, and gross motor equipment. A professional play facilitator guides the child’s play while a team of professionals evaluates the child’s skills. TPBA-2 assesses four domains of development, Cognitive, Communication, Social Emotional, and Motor. In addition, TPBA-2 assesses the development of vision and hearing which were not included in the present study because neither the Bayley-III nor the MSEL assess these areas of development. Each domain is further delineated into seven subcategories. The Cognitive domain consists of attention, memory, problem solving, social cognition, complexity of play, conceptual knowledge, and emerging literacy. The Communication domain consists of receptive language, expressive language, pragmatics, articulation and phonology, voice and fluency, oral mechanism, and hearing. The Sensorimotor domain consists of functions underlying movement, gross motor activity, arm and hand use, motor planning, modulation of sensation, and motor contributions to self care. The Social-Emotional domain consists of emotional expression, emotional style, state and emotional regulation, behavioral regulation, sense of self, social relations, and emotional
themes of play (See Appendix A for detailed definitions of each sub category). The TPBA-2 differs from the original TPBA in the details of the content area rather than in the administration. The subcagagories have been updated to reflect current research, theory, and practice on each of the developmental domains (sensorimotor, emotional and social, communication, and cognitive). The observation process remains the same. The author added forms for obtaining preliminary information from parents and caregivers as well as a hearing and vision screening.

**TPBA-2 Scoring.**

Children’s play skills are observed and age scores are determined for each of the developmental domains. Within each of the subcategories, children’s play skills are listed in a developmental sequence. Age-equivalent (AE) scores from the TPBA-2 are based on the age range guidelines for each subcategory and provide a domain age equivalent score (Linder, 1993). The age equivalent scores are used so that they can be compared to the age equivalent scores of the Bayley-III, MSEL, and Vineland-II. The AE scores can also be converted to standard scores by converting the AE scores into a z score and using the standard formula 15z + 100 to yield a standard score with an average of 100 and standard deviation of 15 (Joseph, Lopez, & Summerall, 2001). This conversion allows comparison of the TPBA-2 with portions of the norm-referenced assessments that do not provide age equivalent scores as well as with the Parent-Rating Scale of Typical Behavior and the evaluator scale of Test Taking Behavior.

**Vineland Adaptive Behavior Scales, Second Edition.**

The Vineland Adaptive Behavior Scales-Second Edition (Vineland-II) is an individually administered measure of adaptive behavior for ages one month through 90.
The Parent/Caregiver Rating Form assesses adaptive behavior in the developmental domains of Communication, Daily Living Skills, Socialization, and Motor Skills. An Adaptive Behavior Composite score is derived from the domain scores. The Adaptive Behavior Composite and domain scores are provided as standard scores with a mean of 100 and a standard deviation of 15 as well as by age equivalents. There is also a Maladaptive Behavior Domain, which was not used in the present study and is not included in the scoring of the Adaptive Behavior Composite.

The Vineland-II reports an overall mean Adaptive Behavior Composite internal consistency reliability coefficient for ages 0-5 as $r = .97$ as well as domain internal consistency reliability coefficients as $r = .92$ (Communication); $r = .89$ (Daily Living Skills); $r = .93$ (Socialization); and $r = .90$ (Motor Skills). The Vineland-II reports an overall Adaptive Behavior Composite test-retest reliability coefficient for ages 0-2 and 3-6 as $r = .96$ for the younger age group and $r = .94$ for the older age group. Mean domain correlations are reported as $r = .90$ for both age groups. The Vineland-II reports an overall Adaptive Behavior Composite inter-rater reliability for ages birth through 6 years as $r = .83$. Mean domain correlations are reported as $r = .78$.

The validity of the Vineland-II has been substantiated for ages one month through 5 years utilizing the General Adaptive Composite of the Adaptive Behavior Assessment System (ABAS-II). The correlation is reported as $r = .70$ indicating moderately high degree of consistency.

*Parent Rating Scale of Typical Behavior.*

The Parent Rating Scale of Typical Behavior is an eight-item questionnaire completed by the parent immediately following the child’s evaluation. Seven questions
reflect the four domains of development, cognitive, language, motor, and social-emotional. The final question is an open-ended response format and is not scored. The respondent is asked to agree on a four-point scale to statements made about his/her child’s typical skill range exhibited at home, school, and the community compared to behavior exhibited during the evaluation. For example, “during this evaluation I saw big movement skills such as walking, hopping, jumping etc. that my child always displays at home; during this evaluation I saw big movement skills such as walking, hopping, jumping etc. that my child often displays at home; during this evaluation I saw big movement skills such as walking, hopping, jumping etc. that my child sometimes displays at home; during this evaluation I never saw any of the big movement skills such as walking, hopping, jumping etc. that my child displays at home.”

A likert scale of one to four was assigned to “no/never, sometimes, often, and always.” There are 28 possible total points with larger scores reflecting a higher parent perception of the evaluation representing his/her child’s typical skill repertoire (See Appendix B).

*Parent Rating Scale of Typical Behavior Test Development.*

The Parent Rating Scale of Typical behavior was created for the current study to ascertain parental perspective on the effectiveness and functionality of the observed assessment. Questions were created through consultation with an expert in the field of early childhood assessment and reviewed for clarity with parents of young children. Following IRB approval in June 2008, cognitive interviews were completed with three parents during a sample assessment in order to insure that the questionnaire served the intent of gathering information about the child’s typical skills and behaviors. The parents
indicated the questions were clear and pertained to the observed assessment. No changes were suggested. Nevertheless, following the data collection phase of the present investigation, it became clear that several parents of children over the age of three had difficulty answering the question related to gross motor skills because the MSEL does not complete a formal evaluation of gross motor skills for children over age three. This question could not be dropped from the final analysis for two reasons. First, not all parents reported difficulty and second, dropping the question would have artificially lowered the score for those children over 3 years-old compared to children under 3 years-old. When the question was not answered, it was treated as a “skipped” question for that individual child.

*Test-Taking Behavior Rating Scale.*

The Test-Taking Behavior Rating Scale is a 12-item questionnaire adapted from the Bayley-II Behavior Rating Scale Record Form (Bayley, 1996) and the Autism Diagnostic Interview-Revised (Rutter, Le Couteur, & Lord, 2003). It is designed to measure the child’s ability to demonstrate adequate attention and arousal, engagement and emotional regulation, as well as the necessary motor and language quality to complete the evaluation. The Test-Taking Behavior Rating Scale is divided by developmental domain with three questions designated to each domain, Cognitive; Language; Motor; and Social Emotional. Questions are rated on a five-point scale indicating the strength of the behavior present during testing. There is a total of 60 possible points with larger numbers reflecting appropriate, non-interfering behaviors during testing. The Test-Taking Behavior Rating Scale is completed by the test administrators immediately following the evaluation (See Appendix C).
Test-Taking Behavior Rating Scale Development.

The Test-Taking Behavior Rating Scale was created for the current study to assess the child’s behavior specifically during the context of a developmental assessment. The twelve questions were formed through a combination of two previously established measures. Nine questions from the Bayley-II Behavior Rating Scale Record Form (Bayley, 1996) capturing Cognitive, Motor, and Social Emotional Development are used and three questions from the Autism Diagnostic Interview-Revised (Lord, Rutter, & Le Couteur, 1994) capturing language development are used. Following IRB approval, cognitive interviews for content analysis were completed with the research assistants during a sample assessment in order to insure that the questionnaire served the intent of gathering information about the child’s test-taking behavior. The research assistants indicated full comprehension during the sample assessment. Inter-rater reliability was also established using 10% of the children assessed following the standardized assessment tool during the current investigation based on the standards of Grayson & Rust, (2001). Cohen’s Kappa coefficients were analyzed to account for chance agreement between raters and evaluated based on the strength of agreement guidelines of Kandis & Polansky, (2003); Landis & Koch, (1977); & Perreault & Leigh, (1989). The Cohen’s kappa for the Test-Taking Behavior Rating Scale was .78, which is considered substantial agreement.

Procedure

Participation recruitment occurred at the family and professional level. Professional and child/family participants in the Transdisciplinary Play-Based
Assessment (TPBA-2) Summer Institute (“Summer Institute”) were asked to volunteer to participate in the current study once they had registered for the institute.

Parents of children participating in the Summer Institute were sent a packet of information that included the following information: 1) a letter explaining the study and requesting participation; 2) an informed consent form (See Appendices D and E); 3) and a pre-addressed and stamped envelope. When parents called to volunteer to participate in the study and/or returned the consent forms, the parent was called to schedule the child for the assessment. A letter was sent to the parent confirming the child’s participation in the study with the dates and time of participation, the assessments location address, and a basic intake form requesting child and family demographics (See Appendix F).

Registered professionals attending the Summer Institute were sent a one-page brochure detailing the study purpose as well as an informed consent form (See Appendices G and H). Registered professionals were asked to submit their professional discipline when registering for the Institute. To control for participant characteristics to ensure internal validity (Gliner & Morgan, 2000), criteria for child and professional participation were established. First, we attempted to recruit children within the age range of 24-47 months; second, participants were English speaking; and third, criteria for professional participation was established. Professionals had to have completed the appropriate degree for practice as mandated by their specific professional organization.

It was not possible however, to recruit a sufficient number of children within the established age range because the evaluations were completed during a weeklong training course for professionals that occurred during a specific time frame. In order to recruit a
substantial amount of children it was necessary to expand the age range of child participants.

TPBA-2 Assessor Characteristics and Recruitment.

Each TPBA-2 team consisted of at least one speech and language pathologist, one occupational therapist or physical therapist, and one educational or school psychologist. Additional team members such as clinical psychologists, social workers, early childhood educators, graduate students or nurses were evenly distributed so that one team consisted either of a social worker or a clinical psychologist but not both. Graduate students were considered additional team members to shadow the professionals and did not comprise a separate team.

Standardized Measures Assessor Characteristics and Recruitment.

To control for examiner effect during the administration of the norm referenced test, two primary research assistants and one secondary research assistant with specific training using the Bayley-III and the MSEL, completed all norm-referenced assessments of children. The primary research assistants completed 90% of the standardized assessments and the secondary research assistant completed 10%. As indicated, each had completed five assessment courses, two general assessment courses, two focused on early childhood, and one focused on social emotional assessment. Additionally, each had participated for a full year at the University Assessment clinic including a proportion completed in the Early Childhood Clinic. Each research assistant had at least three years of graduate school training working with young children.

An effort was made to control for possible order effects by administering the assessments in counterbalanced order. To control for the possible effects of time, the
time of the assessment was administered in counterbalanced order. Children were randomly assigned to two groups, “A” and “B”. Group A children were assessed with the norm-referenced assessment first (either the Bayley-III or the MSEL) and with the TPBA-2 second. Group B children were assessed first with the TPBA-2 and then with the norm referenced assessment. The assessment of children took place over a two-week period in August 2008. Children assigned to group A, were administered the norm-referenced assessment and TPBA-2 during week one. Children assigned to group B, were administered the TPBA-2 during week one and the norm-referenced assessment during week two. All assessments were video recorded for quality assurance using a small digital recorder.

To insure confidentiality, children and professionals were assigned identification numbers to their first name. When the children and parents arrived, the research assistants began following the research protocol (See Appendix I). One research assistant collected the informed consent form if it had not been previously received as well as the child and parent demographic form. The research assistant also took time to build rapport with the child and introduced the Bayley-III or MSEL in a child-friendly manner.

The child and parent were introduced to the playroom at the Fisher Early Learning Center at the University of Denver. To control for the examiner and environmental sources of “construct-irrelevant” (Bracken, 2007) influences, several steps were taken. The playroom consisted of a child size table and two chairs. The room contained inviting colors and materials, but was modestly decorated. To reduce the possibility of distractions, a sign was placed on the door of the assessment room that read, “Do not disturb.” The chair intended for the child during test administration was
placed against a wall so that the chair could not move backwards in order to help the child remain in his/her chair. Finally, the research assistant was asked to avoid wearing bright or stimulating jewelry, clothes, or accessories as well as to limit the amount of jewelry exposed in order to avoid creating visual distractions for the child that might compromise the validity of the assessment.

The administrator assessed the child using the Bayley-III and MSEL according to the manual instructions (Bayley, 2006; Mullen, 1995). Following the assessment, the parent completed the Typical Behavior Rating tool and the Vineland-II. The child was provided a sticker and a book for participation prior to leaving. The administrator explained to the parent that she would contact the parent if she felt like the child would benefit from an additional evaluation and stated she would provide the appropriate referral information at the end of the Institute. The research assistant reminded the parent of the scheduled play assessment prior to the parent’s exit. The administrator scored the norm-referenced assessment as described by the manual. The administrator and the second research assistant completed the Test-Taking Behavior questionnaire for inter-rater reliability analysis on 10% of the children.

During the second portion of the Summer Institute, Group A and B children participated in the TPBA-2. When children and parents arrived, the research assistant explained the purpose and the goal of the study as well as articulated that participation was voluntary and that children and or parents could discontinue participation at any time. The research assistant collected informed consent forms that had not been previously received.
The child and parent were introduced to the playroom. The playroom consisted of the same child size table and chairs as well of a variety of toys, constructive objects, art supplies, and books. The TPBA-2 assessment was completed based on methods described by Linder, (1993, 2008). The team of professionals scored the TPBA-2 in accordance with the methods described by Linder (1993, 2008). The TPBA-2 team completed the Test-Taking Behavior Rating tool. Group B parents completed the Vineland-II. Group A and B parents completed the Typical Behavior Rating tool. The child was provided a sticker and a book for participation prior to leaving. Group A children and parents were provided a $5 gift card to Target and their names were entered into a raffle for a $50 gift certificate to a toy store. The following week, Group B children returned to the Fisher Early Learning Center for the assessment with the norm-referenced assessment. The same procedures were followed as in week one. Group B parents received a $5.00 gift card to Target and were entered into the raffle for the gift certificate.
CHAPTER THREE

Results

The results are based on separate analyses completed for each research question and are explained accordingly. There were no significant differences between the Group A and Group B children. A review of the demographics of the final study sample and presentation of the developmental assessment score conversions and the descriptive statistics of the Transdisciplinary Play Based Assessment-2 (TPBA-2), Bayely Scales of Infant and Toddler Development-III (Bayley-III), and Mullen Scales of Early Learning (MSEL) precedes the detailed explanation of the study results.

Preliminary Results

A total of 39 children were used in the final analyses from the original sample of 40. One child was dropped from the final analyses because her age was beyond the intended age range of the MSEL. The sample size varied for each analysis based on the number of children in the intended age range for the standardized developmental tools, the rate of return on parent completed measures, or child behavior impacting the validity of subtests. As a result, 39 children yielded complete TPBA-2 protocols, 17 out of 19 children yielded complete Bayley-III protocols, and 14 out of 20 children yielded complete MSEL protocols (See Table 3). Analyses using the standardized assessment, TPBA-II, and Vineland-II tools were based on subcategories/subtests as well as a total “Test Score.”
Table 3

*Number of children completing full evaluation tool*

<table>
<thead>
<tr>
<th>Test name</th>
<th>N tested</th>
<th>N able to complete</th>
<th>% of non-completers with identified disability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayley-III&lt;sup&gt;a&lt;/sup&gt;</td>
<td>19</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>MSEL&lt;sup&gt;b&lt;/sup&gt;</td>
<td>20</td>
<td>13</td>
<td>66</td>
</tr>
<tr>
<td>TPBA-2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>39</td>
<td>39</td>
<td>NA</td>
</tr>
</tbody>
</table>

<sup>a</sup> The Bayley Scales of Infant and Toddler Development-III

<sup>b</sup> The Mullen Scales of Early Learning

<sup>c</sup> Transdisciplinary Play Based Assessment-2
First, TPBA-2 domain age equivalents (cognitive, language, social emotional and motor) were converted to standard scores for ease of comparison with the Bayley-III and MSEL. Additionally, a total TPBA-2 Developmental standard score was created by averaging the four TPBA-2 domain standardized scores. Standardized age equivalent scores were created by converting the age equivalent scores to standardized z scores and then applying the formula \(15z+100\). A similar age equivalent score conversion was applied to the Bayley-III and MSEL. For the Bayley-III, the receptive and expressive language and fine and gross motor subscales were averaged to create language composite and motor composite scales. A total Bayley-III Developmental standard score was also created by averaging the four domains (cognitive, language, social emotional and motor).

The structure of the MSEL however, is slightly different because it does not comprise a social emotional scale, the gross motor domain is not used with children over three years old, and it provides a total composite standardized score, the “Early Learning Composite.” The “motor” domain of the MSEL in the presentation of the results therefore represents fine motor development. The Early Learning Composite is the sum of the three developmental domains, Visual Reception, Language, and Motor. The MSEL subscale age equivalents (visual reception, language, motor) were converted to standardized scores. All analyses reflect the comparison between the developmental domains as specified in the manuals for each of the developmental assessment tools. Analyses were not completed to reflect a comparison of content (i.e., subscales were not added to or removed from domains to create content equivalent developmental domains across measures) because in actual practice an individual would not partition an assessment tool to meet arbitrary psychometric properties.
Second, means and standard deviations for each of the measures were calculated. The average total developmental score of the TPBA-2 was 100 (SD = 14.31). The average cognitive score was 100 (SD = 15). The average language score was 100 (SD = 15). The average social emotional score was 100 (SD = 15). The average motor score was 100 (SD = 15). The average Bayley-III total score was 102 (SD = 13.50). The average cognitive score was 100 (SD = 15). The average language score was 100 (SD = 15). The average social emotional score was 107 (SD = 20.01). The average motor score was 100 (SD = 15). With the exception of the Bayley-III social emotional domain in the study sample, the domain averages and standard deviations were similar to the norm samples. The larger average and standard deviation of the study sample Bayley-III social emotional score compared to the other subtest averages and standard deviations on the TPBA-2 and Bayley-III likely reflect demographic differences of the study sample compared to the norm group of the Bayley-III. The study sample could have contained proportionally more children receiving early intervention services, which would teach parents how to be more sensitive in their observations of child behavior and social interactions. This remains only a hypothesis, as the Bayley-III technical manual does not delineate the number of children receiving early intervention services in the norm group.

The MSEL average Early Learning Composite was 98 (SD = 12.90). The average visual reception score was 100 (SD = 15). The average language score was 100 (SD = 15). The average motor score was 100 (SD = 15). For practical comparison, Table 4 demonstrates the averages and standard deviations of the age equivalent scores for the domains of the TPBA-2, Bayley-III, and the MSEL.
Table 4

Child average and standard deviation age-equivalent score for the Transdisciplinary Play Based Assessment-2 (TPBA-2), Bayley Scales of Infant and Toddler Development-III (Bayley-III), and Mullen Scales of Early Learning (MSEL)

<table>
<thead>
<tr>
<th>Test tool</th>
<th>Cognitive</th>
<th>Language</th>
<th>Social emotional</th>
<th>Motor</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPBA-2</td>
<td>29.56 (16.47)</td>
<td>29.82 (17.84)</td>
<td>25.25 (15.59)</td>
<td>29.18 (17.21)</td>
</tr>
<tr>
<td>Bayley-III</td>
<td>20.63 (11.15)</td>
<td>22.84 (13.59)</td>
<td>NA</td>
<td>19.94 (11.68)</td>
</tr>
<tr>
<td>MSEL</td>
<td>49.22 (14.53)</td>
<td>46.69 (16.52)</td>
<td>NA</td>
<td>41.55 (13.67)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are presented in parentheses. There are no age equivalent scores for the Bayley-III or the MSEL Social Emotional Domain.
Secondary Results

The subsequent sections will review the results separately of each research question. The first research question addressed the association between the TPBA-2 and Bayley-III as well as the TPBA-2 and the MSEL. A Pearson product-moment correlation coefficient was used to assess the relationship between the TPBA-2 and the Bayley-III age equivalent developmental scores. An alpha level of .05 was used for all statistical tests. First, Pearson product-moment correlation coefficients were computed within the domains of the TPBA-2 and Bayley-III (intracorrelations). Second, Pearson product-moment correlation coefficients were computed between the TPBA-2 domain age equivalent scores and the Bayley-III domain age equivalent scores as well as between the total developmental TPBA-2 age equivalent score and the total Bayley-III age equivalent score (intercorrelations).

The intracorrelations within the TPBA-2 domains are shown in Table 5. Correlations within the TPBA-2 domains were positive and of high magnitude. The lowest correlations were found between the Social Emotional and Motor Domains (r = .80, n = 39, p < .01). The intracorrelations within the Bayley-III are shown in Table 6. Correlations within the Bayley-III domains were positive and of low to high magnitude. The lowest correlations were found between the Social Emotional Domain and the Cognitive, Language, and Motor Domains (r = .84-.21).

There was a significant positive correlation between the TPBA-2 total developmental score and the Bayley-III total developmental score (r = .95, n = 17, p < .01). The intercorrelations between the TPBA-2 domain scores and the Bayley-III
domain scores are shown in Table 7. Correlations between the TPBA-2 and Bayley-III cognitive, language, and motor domains were all significant at the 0.01 level and of high magnitude. In contrast, correlations between the Bayley-III social emotional domain and all TPBA-2 developmental domains were of low magnitude and not significant.

Table 5

*Intracorrelations within Transdisciplinary Play Based Assessment – 2 (TPBA-2) domains*

<table>
<thead>
<tr>
<th>Domain</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td></td>
<td>.952*</td>
<td>.929*</td>
<td>.863*</td>
</tr>
<tr>
<td>Language</td>
<td>.952*</td>
<td></td>
<td>.907*</td>
<td>.837*</td>
</tr>
<tr>
<td>Motor</td>
<td>.929*</td>
<td>.907*</td>
<td></td>
<td>.801*</td>
</tr>
<tr>
<td>Social Emotional</td>
<td>.863*</td>
<td>.837*</td>
<td>.801*</td>
<td></td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.01 level.*
Table 6

Intracorrelations within The Bayley Scales of Infant and Toddler Development-III (Bayley-III) domains

<table>
<thead>
<tr>
<th>Domain</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cognitive&lt;sup&gt;a&lt;/sup&gt;</td>
<td>___</td>
<td>.965*</td>
<td>.980*</td>
<td>.149</td>
</tr>
<tr>
<td>2. Language&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.965*</td>
<td>___</td>
<td>.957*</td>
<td>.206</td>
</tr>
<tr>
<td>3. Motor&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.980*</td>
<td>.957*</td>
<td>___</td>
<td>.084</td>
</tr>
<tr>
<td>4. Social Emotional&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.149</td>
<td>.206</td>
<td>.084</td>
<td>___</td>
</tr>
</tbody>
</table>

<sup>a</sup> n = 19.  <sup>b</sup> n = 19.  <sup>c</sup> n = 17.  <sup>d</sup> n = 18

*. Correlation is significant at the 0.01 level.
Table 7

*Intercorrelations: Transdisciplinary Play Based Assessment-2 (TPBA-2) and The Bayley Scales of Infant and Toddler Development-III (Bayley-III) developmental domains*

<table>
<thead>
<tr>
<th>Bayley-III</th>
<th>Cognitive</th>
<th>Language</th>
<th>Motor</th>
<th>Social Emotional</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TPBA -2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive</td>
<td>.910*&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.908*&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.923*&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.358&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Language</td>
<td>.970*&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.945*&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.963*&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.152&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Motor</td>
<td>.823*&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.851*&lt;sup&gt;a&lt;/sup&gt;</td>
<td>825*&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.317&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Social</td>
<td>.893*&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.887*&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.906*&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.306&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> n = 19. <sup>b</sup> n = 17. <sup>c</sup> n = 18

*. Correlation is significant at the 0.01 level.
A Pearson product-moment correlation coefficient was again used to assess the relationship between the TPBA-2 and the MSEL age equivalent developmental scores. An alpha level of .05 was used for all statistical tests. Pearson product-moment correlation coefficients were computed within the domains of the MSEL. Pearson product-moment correlation coefficients were then computed between the TPBA-2 domain age equivalent scores and MSEL domain age equivalent scores and total developmental scores.

The intracorrelations within the MSEL are shown in Table 8. Correlations within the MSEL domains were positive and of moderate to high magnitude. The lowest correlations were found between the Language and Visual Reception domains ($r = .86$, $n=16$, $p = .<.01$).
<table>
<thead>
<tr>
<th>Domain</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Visual Reception&lt;sup&gt;a&lt;/sup&gt;</td>
<td>___</td>
<td>.856*</td>
<td>.908*</td>
</tr>
<tr>
<td>2. Language&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.856*</td>
<td>___</td>
<td>.918*</td>
</tr>
<tr>
<td>3. Fine Motor&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.908*</td>
<td>.918*</td>
<td>___</td>
</tr>
</tbody>
</table>

<sup>a</sup>n= 18.  <sup>b</sup>n=16.

* Correlation is significant at the 0.01 level.
As indicated, comparison between the Total Developmental TPBA-2 score and MSEL Early Learning Composite was only completed for 14 of the 20 children who received the MSEL developmental evaluation. Seven children did not have an Early Learning Composite because scores could not be obtained in all three developmental domains (Visual Reception, Language, and Motor) comprising the Early Learning Composite due to the child’s behavior during testing. There was a non-significant positive correlation between the Total Developmental TPBA-2 score and the MSEL Early Learning Composite (r = .12, n = 13, p = .71). The lack of significant relationship may be in part due to the very small sample size.

The intercorrelations between the TPBA-2 domain scores and the MSEL domain scores are shown in Table 9. Correlations between the cognitive, language, and motor domains on the TPBA-2 and MSEL were all significant, positive, and of moderate to high magnitude. The lowest correlations, while still moderate, were between the TPBA-2 Motor Domain and the MSEL Visual Reception Domain (r = .74, n = 18, p < .01) and between the MSEL and TPBA-2 Motor Domains (r = .78, n = 18, p < .01).
Table 9

*Intercorrelations: Transdisciplinary Play Based Assessment-2 (TPBA-2) and Mullen Scales Early Learning (MSEL) developmental domains*

<table>
<thead>
<tr>
<th>MSEL</th>
<th>Visual Reception</th>
<th>Language</th>
<th>Fine Motor</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPBA-2 Cognitive</td>
<td>.788*&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.918*&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.864*&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>TPBA-2 Language</td>
<td>.896*&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.920*&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.856*&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>TPBA-2 Motor</td>
<td>.741*&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.937*&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.784*&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Note. Visual Reception is the Cognitive Domain on the MSEL. There is no MSEL Gross Motor Domain for children over three years old. MSEL fine motor score is compared to the TPBA-2 subcategory “Arm and Hand Use”

<sup>a</sup> n= 18. <sup>b</sup> n = 16

*.Correlation is significant at the 0.01 level.
The second research question explored the ecological validity of the TPBA-2 as measured by the relationship between the TPBA-2 and the Vineland-II on the domains of cognitive, communication, motor, social emotional, and overall development. A Pearson product-moment coefficient was used to assess the relationship between the association between the TPBA-2 and daily functional skills as measured through the Vineland -II. The results of the Pearson product-moment coefficients are presented based on 33 children, however where parents skipped questions, domain scores could not be calculated and this is reflected in the different sample sizes. Six parents either did not complete or did not return the Vineland-II questionnaire form on their child. On the cognitive domain, a comparison was made between the TPBA-2 cognitive domain and the Vineland-II daily living skills domain, which assesses problem solving, conceptual, and the mental manipulation required for daily cognitive skills.

The correlation between the TPBA-2 overall developmental score and the overall Vineland Adaptive Behavior scale score was not significant \((r = .24, n = 28, p = .22)\) suggesting there is no relationship between the created overall TPBA-2 developmental score and the total Adaptive Behavior scale of the Vineland-II.

In contrast to the total score comparison, the correlations between the sub tests were positive and significant. The correlation between the TPBA-2 cognitive domain and the Vineland-II daily living skills was positive and significant \((r = .87, n = 32, p < .01)\) suggesting a high degree of overlap between the measured cognitive skills on the Vineland-II and TPBA-2. The correlation between the TPBA-2 language domain and the Vineland-II communication domain was positive and significant \((r = .83, n = 32, p < .01)\)
suggesting a high degree of overlap between the measured language skills on the Vineland-II and TPBA-2. The correlation between the TPBA-2 motor domain and the Vineland-II motor domain was positive and significant ($r = .69$, $n = 30$, $p < .01$) suggesting a moderate degree of overlap between the measured motor skills on the Vineland-II and TPBA-2. This slightly lower correlation likely reflects the primary focus of the Vineland-II Motor section on gross and fine motor skills, whereas the TPBA-2 has an additional focus on sensory processing and motor planning. The correlation between the TPBA-2 social emotional domain and the Vineland-II social emotional domain was positive and significant ($r = .82$, $n = 33$, $p < .01$) suggesting a high degree of overlap between the measured social emotional skills on the Vineland-II and TPBA-2.

The third research question explored possible differences between the ability of the TPBA-2 and the standardized assessments to demonstrate the child’s most typical skills and abilities during the assessment. Results are presented based on a comparison between the TPBA-2 and the combined standardized developmental assessment score on the Parenting Rating Scale of Typical Behavior. A Paired-Sample t-test was used to compare the average typical behavior scores. There were 30 complete pairs. Raw scores on the Parent Rating Scale of Typical Behavior were first converted to standardized scores for ease of understanding. There was a slight trend towards more observed typical child behavior during a play-based assessment however, there was no significant difference in the caretaker report of perceived demonstration of the child’s typical behavior during a play assessment ($M = 99.78$, $SD = 15.44$) and during a standardized developmental assessment ($M = 98.83$, $SD = 15.27$), $t(29) = .25$, $p = .808$ (two-tailed).
The final research question investigated possible differences in the evaluator’s perceived child’s test taking behavior during the play assessment and the standardized developmental assessments. Results are first presented describing the relationship between the child’s behavior during testing and subsequent test score and are presented separately for the TPBA-2, Bayley-III and MSEL. Pearson moment product coefficients were used to evaluate the relationship between the evaluator’s perception of the child’s behavior during testing and the child’s test score. There was a significant relationship between test taking behavior score, as measured by the Test Taking Behavior Rating Tool, and the total TPBA-2 developmental score (r = .40, n =38, p =.001). There was a significant relationship between the test taking behavior score and the total Bayley-III developmental score (r =.77, n =16, p =.001). In contrast, there was no significant relationship between the child’s test taking behavior score and the MSEL Early Learning Composite (ELC) score (r =.22, n =13, p =.46) which may be the result of a small sample of children with a complete ELC score.

The final analysis explored the potential differences in the evaluator’s perceived child’s test taking behavior during the play assessment and the standardized developmental assessment. The results are presented based on a comparison between the TPBA-2 and the combined standardized developmental assessment score on the Test Taking Behavior Rating Tool. A Paired-Sample t-test was used to compare the average test taking behavior scores. There were 38 complete pairs. Raw scores on the Test Taking Behavior Rating Tool were first converted to standardized scores for ease of understanding.
A child’s behavior during the evaluation did not appear to be different, as perceived by the evaluator, depending on the type of evaluation the child received. There were no significant differences between the test taking behavior score of children during the play assessment (M = 100, SD = 15) and the standardized developmental assessments (M = 99.85, SD = 15.17), t (37) = .08, p = .932 (two-tailed) as perceived by the evaluator.
CHAPTER FOUR

Discussion

The final chapter will discuss the results in detail of each research question. The discussion begins with a brief summary of the findings and an explanation of the intracorrelations yielded within the TPBA-2, Bayley-III, and MSEL and the intercorrelations between the TPBA-2 and Bayley-III and the TPBA-2 and MSEL. The results and implications of each research question are then presented separately and the chapter concludes with a discussion on the limitations of the current investigation, future directions, and clinical implications.

Previous evaluations of the Transdisciplinary Play Based Assessment focused on the first version of TPBA, and demonstrated concurrent and content validity as well as the test-retest reliability (Cornett, 1998; Friedli, 1994). The current study was the first investigation of the updated version of the TPBA, the TPBA-2, compared to the most updated and available standardized developmental instruments of the Bayley-III and the Mullen Scales of Early Learning. It was the first study to empirically explore the construct of ecological validity within the context of this early childhood assessment. This study was also the first to investigate the impact of the assessment tool (standardized or play-based) on the facilitation of the child’s typical set of skills. Finally, the current study was the first to empirically explore the relationship between young children’s test taking behavior and test score.
The results of the current study demonstrated the concurrent validity of the TPBA-2, using all developmental domains and the team approach outlined by Linder, 1993, 2008, with the Bayley-III and the MSEL. Findings suggest the TPBA-2 also demonstrates ecological validity, successfully revealing information about what a child demonstrates in his/her typical environments (Chaytor & Achmerrer-Edgecomb, 2003). Inconclusive and non-significant results were found regarding the differences in parents’ perceptions on whether the type of assessment tool (play or standardized) better captured a child’s typical repertoire of behaviors that are observed at home, school, and the community. Inclusive and non-significant results were also found regarding the relationship between a child’s behavior during testing and subsequent test score as well as the differences in the child’s behavior during the testing as perceived by the evaluator. A detailed discussion of the results is presented in the following sections.

*Bayley-III, MSEL, TPBA-2 Inter and Intracorrelations*

In the current investigation, observed intra correlations (i.e., the correlations found among the domains within the Bayley-III, MSEL and TPBA-2) were higher than the guidelines suggested by the American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education. The observed intra correlations within the Bayley-III and MSEL were also higher than the correlations in the published norm groups (Bayley, 2006; Mullen, 1996). Similar to the intracorrelations, the intercorrelations (i.e., the correlations observed between the TPBA-2 and the standardized assessment tools) were higher than the stated guidelines however, the observed magnitude of correlations between the TPBA-2 and MSEL are comparable to previous research comparing a play assessment (Play Assessment Scale) with the
MSEL (Esiert & Lamorey, 1996). Due to the observed high intra and intercorrelations, a brief discussion is warranted.

The intracorrelations of the Bayley-III subtests for the norm group across all ages are reported to range from .21 (social emotional with motor) to .52 (expressive language with cognitive) (Bayley, 2006). The intracorrelations for the MSEL norm group are reported to range from .46 (visual reception with expressive language) to .67 (receptive language with expressive language) (Mullen, 1996). In contrast, the observed intra correlations in the current study for the Bayley-III ranged from .08 (social emotional with motor) to .980 (cognitive with motor) and .86 (visual reception with expressive language) to .92 (fine motor with receptive language) for the MSEL.

The magnitude of the observed intra and inter correlations is likely due to the sample size and demographics of the study sample. Two significant factors impacting the Pearson product-moment coefficient are sample size and sample variability (Coladarci, Cobb, Minium, & Clark, 2004). Small sample sizes are more likely to create larger correlations than what would be found in the population (David 1954, Hays, 1963 as cited by Kareev, Lieberman, & Leve, 1997). Additionally, while a restricted range of variability would generally lower correlations, a diverse sample and score range could help increase the correlation coefficient (Coladarci, Cobb, Minium, & Clark, 2004). In the current investigation, the sample consisted of families from a range of socioeconomic backgrounds and a range of child ages and developmental levels with 51% of the children identified with a developmental delay or disability as compared to 10% of the norm sample in the Bayley-III and 0% of the norm sample in the MSEL (Bayley, 2006; Mullen, 1996). While the diversity of the study sample facilitates generalized
conclusions, the diversity of the study sample within the context of a smaller sample size, may have inflated the correlation coefficients for both the correlations within each assessment tool and between the assessment tools.

Although sample size and variability impact correlation coefficients, moderate to high correlations within the TPBA-2 domains may be additionally theoretically grounded. Specifically, TPBA-2 assumes interdependence among developmental domains and that developmental growth in one area closely precedes, parallels or influences developmental growth in another area (Linder, 2008). This assumption, which has been previously empirically substantiated, (See Linder, Goldberg, & Goldberg, n.d) presumes a close and even interconnected relationship between developmental domains and would as a result increase correlation coefficients within the developmental domains of the TPBA-2. The following sections explore in detail the results of each research question.

*Concurrent Validity: Is there a statistically significant association between the TPBA-2 and Bayley-III?*

The results demonstrated concurrent validity between the TPBA-2 and two widely used standardized developmental inventories when using age equivalent scores. There were significant associations between the TPBA-2 cognitive, language, motor and the corresponding Bayley-III domains. Additionally, a significant association was demonstrated between an overall TPBA-2 developmental age equivalent score and an overall Bayley-III developmental age equivalent score, though in actual practice, a total TPBA-2 developmental score would not be generated or appropriate, as the differences across domains and subcategories are considered more relevant (Linder, 2008). A non-
significant relationship was observed between the social emotional domain of the Bayley-III and all developmental domains of the TPBA-2. The potential relationship between the Bayley-III and the TPBA-2 social emotional domains is the most salient in the exploration of the concurrent validity of the TPBA-2 social emotional domain.

The non-significant relationship between the two social emotional domains may reflect deviations in measurement specific to the study sample, which are, in turn, magnified by the brevity of the Bayley-III social emotional questionnaire. Differences between the two measures in how information is collected about the child’s social emotional development may also yield very different results across the two measures.

As reported, the standard deviation of the Bayley-III social emotional domain was larger than the standard deviations of the other Bayley-III and TPBA-2 domains. Since the product of the standard deviations from the two variables of interest (i.e., the Bayley-III and TPBA-2 social emotional domains) comprise the denominator of the correlation coefficient, a larger standard deviation of one variable would decrease the value of the correlation coefficient between the Bayley-III and TPBA-2 social emotional domains. Additionally, the Bayley-III social-emotional questionnaire was originally designed by Greenspan (Greenspan, Greenspan Social-Emotional Growth Chart, 2004) as a screening questionnaire and was subsequently used as a subtest for the Bayley-III. It contains a total of only 35 items, of which parents complete only a portion depending on the child’s age. There are very few items in each sub category and coupled with the small sample size of the current study, the analysis is limited to very few data points to compare with the TPBA-2 social emotional domain, which includes numerous items across seven sub-categories.
Alternatively, the non-significant relationship could reflect the different methods of data collection about social development between the TPBA-2 and the Bayley-III. The Bayley-III method is uniquely through parental report. In contrast, the TPBA-2 gathers data through direct clinical assessment and is also supplemented with parental report. By incorporating multiple data sources of clinical observation and parental report, the inherent process of the TPBA-2 has a “checks and balance” system against parental bias, (i.e., over or under reporting behavior problems or competencies. (See Carter, Briggs-Gowan, & Ornstein Davis, 2004 for review), and/or test environment bias (i.e., the novelty of the test environment suppressing low-base rate behaviors). It is interesting to note that the standard scores on the social emotional domain on the Bayley-III were significantly higher than the converted standardized age equivalent social emotional scores on the TPBA-2. The reasons for this difference could be the result of several factors, including but not limited to parental bias or the professional skill needed to perform a TPBA-2 (Bracken, 1988), and are beyond the scope of the current investigation.

There are also theoretical differences related to construct between the two measures of social emotional development. Both the Bayley-III and TPBA-2 social emotional subtests share a common philosophy that cognitive, language, motor, and social emotional development, is a dynamic and fluid system where each can influence or is influenced by the other. The items on the Bayley-III social emotional subtest reflect this theoretical assumption within one domain, whereas; items within the TPBA-2 are unique to the given developmental domain. Some of the content of the Bayley-III social emotional items such as joint attention, sensory regulation, functional language, social
reciprocity, and early motor development, while crucial for adequate social emotional functioning, are not direct measures of social and emotional development and are usually assumed under other development realms, namely language and motor. In contrast, the content of the seven sub-categories of the TPBA-2 social emotional development reflect content areas exclusively assumed under social emotional development, such as emotional expression, emotional/behavior regulation, development of self, emotional themes in play, and social interactions.

A definitive conclusion about the concurrent validity of the TPBA-2 social emotional domain cannot be made based on the current results due to problems with the varying methods of data collection, and the differences in the goals of each measure. There will, however, be challenges to future investigations. Compared to the cognitive, language, and motor domains of development, social emotional development has been characteristically more difficult to evaluate because there are not easily definable narrow behaviors to assess at each age (Cater, Briggs-Gowan, & Orstein Davis, 2004; Losardo & Notari-Syverson, 2001). It is difficult to narrowly define a set of observations that takes into account individual child differences, variations in culture and socialization practices, and the normative behavior and developmental challenges that occur in early childhood development (Owens, 2007). Current tools of social emotional development tend to use parent report, do not always include a validity scale, are focused on pathological behaviors, or are primarily used as screening tools. Most social-emotional assessment instruments do not cover a similar breadth and depth of social emotional development as the social emotional domain of the TPBA-2, which makes comparisons complicated.
Concurrent Validity: Is there a statistically significant association between the TPBA-2 and MSEL?

A strong association was observed between the TPBA-2 cognitive, language and motor domains with the corresponding MSEL developmental domains (visual reception, language, and motor). A significant association was not demonstrated between the overall TPBA-2 developmental age equivalent and the Early Learning Composite (ELC) of the MSEL.

There are several explanations to the non-significant relationship between the ELC and the TPBA-2 total developmental score. First, there was a very small sample of children (i.e., 13) who were able to obtain a valid profile to produce the ELC. There were seven children who could not complete the MSEL to produce a valid ELC score due to behavioral difficulties or because of an inappropriate match between the test content and structure and the child’s developmental needs. In contrast, all children were able to complete the TPBA-2, perhaps indicating the more stressful nature of the structured tasks on the MSEL. The comparison incorporated only the 13 children who had a complete ELC score. The reduced sample likely had inadequate power to yield significant results (Cohen, 1992). A second possibility is that the ELC may represent a different construct than the TPBA-2 overall developmental age equivalent score because the ELC does not incorporate either the gross motor or social emotional domains. The two general developmental scores therefore, may not be comparable.

Ecological Validity: Does the TPBA-2 demonstrate ecological validity as measured by the association between the TPBA-2 and Vineland-II on the domains of cognitive, communication, motor, social emotional, and overall development?
The TPBA-2 cognitive, communication, motor, and social emotional domains appear to demonstrate ecological validity as compared to the Vineland-II, an established ecologically valid measure designed to assess a child’s daily functioning across developmental domains. This finding suggests the TPBA-2 is also an adequate measure to capture everyday, functional capacities, corresponding to what a child does in real life performances necessary in the child’s home, school, and community (Chaytor & Schmitter-Edgecombe, 2003). This finding helps to provide empirical evidence for the ecological validity of TPBA-2, a newer form of validity that has been previously neglected in the literature (Barnett & Macmann, 1992).

In contrast to the domains, a significant relationship was not found between the Total TPBA-2 developmental score and the Adaptive Behavior Score/Composite. This is a peculiar finding given the moderate to high correlations found between the domains. There is not a clear explanation to this non-significant finding and it may simply be a result of the smaller sample of children obtaining a full Vineland Adaptive Behavior score (n = 28) compared to the sample of children able to receive subtest scores (range = 30-33). Alternatively, some items on the sub domains of the Vineland-II are not included in the total score and this may have altered the total score. The comparison however is mostly academic because in practice, a total TPBA-2 developmental score would not be generated because TPBA-2 looks at the range of abilities within and across domains rather than determining one “score” to illustrate the child’s functioning. Additionally, there is greater utility in making comparisons between the domains of TPBA-2 and Vineland-II because they provide more specific information representing the developmental categories targeted for intervention.
Is there a statistically significant difference between the ability of the TPBA-2 and the standardized developmental assessments to demonstrate the child’s most typical skills during the assessment as measured by parent evaluation of the child’s typical home and community behaviors on the Parent Rating Scale of Typical Behavior?

The current investigation found a slight trend suggesting the TPBA-2 may better facilitate a child’s typical behavior and skills during an evaluation. A significant difference was not, however, demonstrated regarding the ability of the type of assessment (i.e., standardized or play-based) to elicit the child’s most typical repertoire of skills during the evaluation as perceived by parents. The non-significant finding may be truly non-significant, though it could also be the result of measurement error. The Parent Rating Scale of Typical Behavior, completed by parents, apparently did not adequately address what it intended to measure despite efforts to obtain content validity. Written comments from parents during the investigation suggested some parents perceived the measure as a tool to evaluate the individuals or team completing the developmental assessment rather than as a tool to assess how the developmental assessment facilitated skills the parent typically observes at home, school, or the community. The non-significant finding likely reflects the confusion parents experienced while completing the questionnaire. Due to problems with the measurement instrument, this study was unable to accurately determine if one approach was better able to elicit the child’s typical skill repertoire, including infrequent behaviors that may only be observed in specific contexts. The findings in the current investigation contrast a previous investigation suggesting that
how successful the assessment tool is at facilitating typical skills can influence test scores. deBruin (2005) found a significant relationship between the original TPBA developmental scores and the Battelle Developmental Inventory scores only when parents felt the child demonstrated his typical behavior during both assessments.

Is there a significant correlation between test score and child test taking behavior as measured by the relationship between the developmental assessments and the Test Taking Behavior Rating tool?

The current study investigated the relationship between the child’s test score and the child’s behavior during testing. A child’s behavior during testing was defined as adequate attention and arousal, engagement and emotional regulation, as well as the necessary motor and language quality to complete the evaluation. It also investigated whether or not a difference existed in the child’s behavior depending on the type of assessment received. A significant positive relationship was observed between a child’s test taking behavior score during the administration of the Bayley-III and during the TPBA-2 suggesting there may be a relationship between a child’s test score and his behavior during test administration (as perceived by the evaluator), which is consistent with previous hypotheses about the negative impact of young children’s behavior on subsequent test scores (Kelly-Vance, Needelman, Troia, & Oliver Ryalls, 1999). A non-significant relationship was observed between a child’s test taking behavior score and subsequent standardized score during the administration of the MSEL. There were no differences found between the quality of the child’s behavior impacting the child’s score during a play assessment or a standardized developmental assessment.
Similar to the Parent Rating Scale of Typical Behavior, the Test Taking Behavior Rating Tool elicited uncertainty on the part of the standardized assessment administrators and the play-based assessment evaluators. All evaluators reported that the measure was confusing to use during the actual investigation and indicated it was not clear whether the child’s behavior during the evaluation should be compared to a perceived understanding of a “normal/typical child” or whether the child’s behavior should be evaluated based on the child’s own performance during the developmental evaluation. As a result, the tool likely did not consistently address what it intended to measure, which was the child’s average behavior, in reference to himself, during the testing session.

The significant positive relationship between the child’s behavior score during the administration of the Bayley-III and TPBA-2 and subsequent developmental score may reflect the evaluator’s perception of the child’s ability to maintain adequate attention and arousal, engagement and emotional regulation, as well as the child’s ability to demonstrate adequate motor and language quality, resulting in a more valid estimate of the child’s abilities. This significant relationship, however, is tempered by the non-significant relationship found between the child’s behavior during testing and subsequent MSEL developmental score as well as the non-significant difference found between the type of developmental assessment condition (i.e., play or standardized) and the quality of the child’s behavior during testing. While the non-significant relationship between the child’s behavior score during the administration of the MSEL and the subsequent developmental score could simply reflect the small sample size of children completing the MSEL, both the significant and non-significant findings are more likely related to measurement error and chance. In this context, the significant association between test
taking behavior and the child’s developmental score on the Bayley-III and the TPBA-2 are more likely reflective of chance than a true significant finding. As a result, due to problems with the measurement instrument, this study was unable to accurately determine if there is a significant relationship between a young child’s test behavior and test score as well as whether there is a difference in a child’s behavior and resulting test score based on the type of assessment received.

*Other findings*

Additional findings outside of the results related to the research questions are worth noting. More than half of the study sample consisted of children with developmental disabilities ranging from minor language delays to more involved disabilities such as Autism Spectrum Disorder and dual diagnosis. All 40 children were able to receive a full Transdisciplinary Play Based Assessment without alternating the assessment process, whereas only 13 of the 21 children who received the Mullen Scales of Early Learning were able to complete the full assessment to obtain a valid profile. The children who did not obtain a complete MSEL had been identified with significant developmental delays. These results echo the findings of Bagnatto and Neisworth’s survey of professionals completing standardized assessments with preschool children. Professionals in the survey reported untestable data for over half of the children assessed when using a standardized tool or even the accommodations provided in the protocol for children with disabilities (Bagnatto & Neisworth, 1994). In the survey, professionals reported using play assessment to obtain information about the child’s development. The results of the current study provide evidence that a play assessment, specifically the Transdisciplinary
Play Assessment, can naturally accommodate the needs of all children to obtain an accurate and developmental profile.

Professionals on the research teams completing the Transdisciplinary Play Based Assessment-2 also provided informal positive feedback related to their experiences using TPBA-2 with children and families. Professionals on the TPBA-2 teams stated that parents of children they assessed reported the information received about their child was “valuable” “detailed” and “extremely useful.” Comparable feedback was not reported by the professionals completing the standardized testing. One team member indicated that her experience learning and completing the Transdisciplinary Play Based Assessment-2 opened her “eyes to how important parents are to the process of assessment.” Parents are crucial to the assessment process but their participation and voice can be lost during typical evaluation environments where the child is alone with the evaluator or time is constrained.

Professionals evaluating young children will ultimately have a choice between two valid methods of assessing young children- a traditional standardized measure which yields a standardize score or a Transdisciplinary Play Based Assessment-2 which yields a developmental age level. On the one hand, the standardized measure may seem faster and less expensive because there is only one individual completing the evaluation rather than an entire team of professionals. Parents may find it more appealing or comforting to receive a standardized score where they can compare the functioning of their child to same aged peers. However, if the child has to be assessed by each member of a team on a different standardized test, families are often required to return several times for cumulative testing. Additionally, if the child has already been identified with a
developmental disability, or is receiving an evaluation to identify a possible disability or delay, a standardized tool may ultimately take more time and may even provide inaccurate information about the child because the standardized tool may not have appropriate norms (or no norms) for children with disabilities. The individual completing the evaluation may begin one test, be able to complete only portions of the test, and need to search for other tools to fill in the gaps of information. The child may then have to be referred on for additional testing to obtain an accurate picture of his/her specific developmental needs in areas not previously addressed, such as language or motor. In contrast, the same child receiving a Transdisciplinary Play Based Assessment-2 would receive a full evaluation assessing all developmental domains during one evaluation. No child is called “untestable.”

Additionally, the administration of the TPBA-2, like a single subject study, provides assessment results with direct implications for intervention over the course of the evaluation. As such, the TPBA process resembles the principles of Therapeutic Assessment (Finn, 2007), resulting in an active and fluid assessment experience where parents are crucial participants in the evaluation. As a result parents learn not only the information provided by the assessment, but also about the interventions used by the professionals to help illicit the child’s full range of skills during the course of the evaluation. Thus, the benefits of the TPBA-2 go beyond identification of disability and ultimately may exceed the financial and time costs to administer it.

Limitations

There are several limitations to the current study, which compromised both the internal and external validity. Sample size was problematic. While we attempted to
recruit children between two and three years old, this proved to be difficult in the allocated time frame particularly because it was important that the sample include both typically developing children and children with an identified developmental delay or disability. As a result, the age range was too large which made the sample size of each age grouping small and may have inflated the magnitude of the correlations when they were significant and decreased the likelihood of finding significant relationships when there may have been one.

A second compromise to internal validity was the two measures created specifically for the investigation: The Test Taking Behavior Rating Tool and The Parent Rating Scale of Typical Behavior. There was not adequate validation of the two measures prior to the investigation resulting in measurement error. Several steps were, however, taken to obtain validity and reliability. Specifically, The Test Taking Behavior Rating Tool was a previously used behavioral observation tool from the second edition of the Bayley with three added questions from a well known and validated diagnostic interview, The Autism Diagnostic Interview (Lord, Rutter, & Le Couteur, 1994). Additionally, the trained research assistants co-observed 10% of the sample of children to obtain reliability. In regards to the Parent Rating Scale of Typical Behavior, content validity had been established through interviews with parents following completion of three sample assessments. Nevertheless, these efforts were likely not sufficient. Measurement error may have also been introduced during the process of completing the forms. Parents were asked to complete the Parent Rating Scale of Typical Behavior immediately following the evaluation while simultaneously having to care for their child.

There were also threats to external validity. The test environments during the
standardized assessment and play assessment, while conforming to the standardized test manuals as well as to the guidelines outlined by Bracken (1988, 2007) to reduce external distractions, were not necessarily similar to the demands of a “real” testing environment, where evaluators have to contend with challenges such as limited space, time, or inappropriate testing rooms. The restrictions of a research study provide optimal testing conditions that are not likely to occur in a “real world” application which may have resulted in an inflation of similarities between the standardized and play assessment. Additionally, the TPBA-2 teams consisted of practicing professionals, previously untrained in TPBA-2, while the standardized test evaluators, though graduate students, had been previously trained and had practice with the tools.

Finally, although the current investigation demonstrated the concurrent and ecological validity of the TPBA-2, it is well known that concurrent validity is not an “entirely satisfactory aspect of validity” (Crocker & Algina, 1986, p. 5). There is clearly a need for further evidence of validity including an examination of what the TPBA-2 cannot measure; its ability to discriminate children who require early intervention from children who are developing within normal limits of development; as well as its ability to predict a child’s future developmental trajectory.

**Future Directions**

In addition to the exploration of the other facets of validity (discriminant, convergent, and predictive validity), exploration of the validity of the Social Emotional Domain of the TPBA-2 is crucial because the results of the current study were inconclusive. There are few existing validated instruments assessing the social emotional development of young children, which makes validating the social emotional domain of
the TPBA-2 challenging. Existing measures focus on problem behaviors such as the Child Behavior Checklist (CBCL: Achenbach & Rescorla, 2000); have a combined focus of problem behaviors and competencies, such as the Infant Toddler and Brief Infant Toddler Social Emotional Assessment (ITSEA/BITSEA: Briggs Gowan & Carter, 2006); or are more brief and considered a screening tool such as the Ages and Stages Questionnaire (ASQ: Squire, Bricker, Twombly, Yockelson, Schoen, & Younghee, 2003) and the Greenspan Social Emotional Growth Chart (Greenspan, 2004). More work is needed in the focused assessment of early childhood social emotional development across a broad age range, birth to 3 years old. The TPBA-2 might be the pioneer in this endeavor.

Future investigations will need to focus on the reliability of the TPBA-2, particularly, test-retest reliability, to document the consistency over time of the TPBA-2 evaluation. Future investigations would benefit from a larger, stratified sample, to determine whether specific populations of children (i.e., children with autism, Down syndrome, speech delay, etc.) are more or less served from use of a play-based evaluation compared to a standardized assessment.

It would also be helpful to understand whether the validity and reliability of the TPBA-2 is dependent upon the team. In the current evaluation, the team consisted of a wide range of professional experiences, there were no teams with professionals with less than three years of professional experience though none had previous training in TPBA-2. It may be that it is necessary to have a well-seasoned team to produce a TPBA-2 developmental evaluation that is comparable to a standardized evaluation. For example, the deBruin (2005) investigation, comparing the original TPBA and the Battelle
Developmental Inventory, employed TPBA teams consisting of a preschool teacher, a paraprofessional, a speech and language pathologist, and a parent, where as the evaluators completing the Battelle Developmental Inventory were licensed specialists in school psychology. deBruin (2005) did not demonstrate congruence between the original TPBA and the Battelle Developmental Inventory unless parents felt both tests demonstrated their child’s typical behaviors.

Future investigations should consider evaluating the impact of the process of the TPBA-2 on parents’ understanding of their child’s development and early intervention systems because in the current study, only scores were compared. The TPBA-2 process however is more complex, including parents as part of the team, before, during and after the assessment in a way that may be perceived as different (and more helpful) from a standardized developmental evaluation. Additionally, examining the resulting reports across approaches would also contribute to understanding the ecological validity of various approaches.

Finally, professionals in early childhood development and assessment have written extensively on the benefits of play as a method of direct observation of a child’s development (Esert & Lamorey, 1996). Previous research with the original TPBA suggested professionals involved in the play assessment process had a greater positive perception of the assessment and of the written reports in regards to its ability to effectively meet the standards of IDEA compared to standardized measures of early child development (Myers, McBride, & Peterson, 1996). Despite professional enthusiasm and empirical support of play as a viable assessment tool, there continues to be a gap in proposed best practices and actual clinical practices. Work is needed to investigate
perceived obstacles of using the TPBA-2, as well as the necessary steps to begin incorporating play-based assessments into routine, funded, developmental evaluations.

*Clinical Implications*

The current findings confirm it is possible to utilize a play-based assessment, particularly the Transdisciplinary Play Based Assessment-2, in lieu of a conventional, standardized assessment tool to determine the age level, often identified as the “adaptive level” of a child. While experts in early childhood development and assessment have recognized the limitations to using standardized assessments with young children, especially children with special needs, (Bagnatto & Neisworth, 1994) and have expressed professional interest in the value of a play-based assessment (Eisert & Lamorey, 1996), there has previously been limited research documenting the efficacy of a play based assessment and its ability to identify a child’s adaptive level comparable to a conventional developmental assessment.

This was the first study to compare the Bayley-III, Mullen Scales of Early Learning, and the Vineland Adaptive Behavior Scales-II, three widely used traditional standardized assessment tools, with the most updated play assessment, the TPBA-2, using both typically developing children and children with special needs. It was also the first investigation to empirically evaluate the function of TPBA-2 as the author intended, which is to employ a professional team working together within and across disciplines to create a transdisciplinary assessment of the child’s development.

The findings underscore the strength and utility of play to produce meaningful information about a child across all domains of development- cognitive, language, motor, and social emotional- to inform functional recommendations. The results empirically
validate the TPBA-2 as a “developmentally appropriate” (Bredekamp, 1997; Neisworth & Bagnato, 2000) early childhood assessment tool and one that meets the standards as outlined by the Individuals with Disabilities and Education Act (IDEA) for choosing an assessment tool for young children. TPBA-2 demonstrates concurrent and ecological validity within the child’s authentic context of play, it incorporates two critical elements, parents and or family member involvement and the accommodation of the child’s unique development and/or disability, and it is multisourced and completed by a multidisciplinary team (Fewell, 1991; Harrington & Tongier, 1993; Neisworth & Bagnato, 2000).

The findings also suggest that TPBA-2 successfully conforms to the six general principles and recommendations established by the Goal 1 Guidelines by the United State’s Congress including validity, age appropriate content and methods of data collection, linguistically and culturally appropriate content and procedures, and the inclusion of parents as a valid source of assessment information and audience for the assessment results (Goals Panel, as cited by Shepard, Kagan, & Wurtz, 1998).

The findings also have practical implications. Professionals have voiced concern regarding the ability of the items on standardized assessment materials to address the developmental skills of children across cultures (Liao, Wang, Yao, & Lee, 2005). The current results and previous research on play assessment suggest that purposeful, strategic observation of play can take less time (Myers, McBride, & Peterson, 1996), utilize flexibility in materials chosen for the assessment, be completed in any setting (Linder, 1993), and adapt to the culture and unique needs of the child and his/her family (Uys, Alant, & Lloyd, 2005). As a result, it may be appropriate to assess children using the
TPBA-2, where the cultural, linguistic, or developmental needs of the child require a flexible but meaningful approach or where access to early assessment and intervention services are limited. For example, the play assessment could readily be completed through video conferencing where a professional team observes via live video stream the play facilitator assessing the child. Immediate and direct parental feedback could similarly be provided, resulting in a transdisciplinary assessment of a child who might not have had the opportunity to receive even a single-disciplined evaluation due to the lack of access to professionals or to funding. Video streaming of live Transdisciplinary Play Based Assessments has already been attempted with notable success (T. Linder personal communication, March 10, 2009), though there is not yet empirical support of the efficacy.

Finally, results of this investigation suggest that a strengths based approach, using multiple data sources of parent report and purposeful child observation, within the child’s natural environment of play, can give professionals and parents accurate developmental information that corresponds to the information obtained from standardized subtests and provides both qualitative and skill level information to inform intervention recommendations (Neisworth & Bagnato, 2000).
References


### Appendix A

**Definition of Transdisciplinary Play-Based Assessment Sub-Categories**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Sub-domain</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention</td>
<td></td>
<td>Ability to select stimuli, focus on the stimuli, sustain concentration, shift focus and ignore distractions.</td>
</tr>
<tr>
<td>Memory</td>
<td></td>
<td>Ability to recognize, recall, or reconstruct routines, skills, concepts, and events after short term and/or long term delays.</td>
</tr>
<tr>
<td>Problem-solving</td>
<td></td>
<td>Ability to understand causal relationships and to independently organize and sequence thoughts and actions toward a goal in a timely process, to monitor progress, to make modifications as needed, and to generalize what is learned to new situations.</td>
</tr>
<tr>
<td>Social Cognition</td>
<td></td>
<td>Ability to infer social causes of events, to understand the thinking and intentions of</td>
</tr>
<tr>
<td>Domain</td>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Complexity of Play</td>
<td>The highest level and predominant type of play exhibited by the child, including sensory motor play, physical play, functional/relational play, construction, dramatic play and games with rules.</td>
<td></td>
</tr>
<tr>
<td>Conceptual Knowledge</td>
<td>Ability to recognize or recall personal or conceptual information related to people, objects, events, categories, characteristics, and pre-academic concepts.</td>
<td></td>
</tr>
<tr>
<td>Emerging Literacy</td>
<td>Understanding and use of books; picture and story comprehension; story reading behaviors, phonemic awareness, letter recognition, word recognition; drawing and writing.</td>
<td></td>
</tr>
<tr>
<td>Domain</td>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Communication</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receptive Language</td>
<td>Ability to understand, process, and respond to verbal and nonverbal language including vocabulary and concept words, questions, grammatical structures, and requests.</td>
<td></td>
</tr>
<tr>
<td>Expressive Language</td>
<td>Ability to use nonverbal and verbal language to express thoughts and feelings, relate to events, and answer and ask questions.</td>
<td></td>
</tr>
<tr>
<td>Pragmatics</td>
<td>Ability to use nonverbal and verbal communication in different social contexts including responding to language, taking turns, initiating interactions, exchanging information, and maintaining a topic.</td>
<td></td>
</tr>
<tr>
<td>Articulation/Phonology</td>
<td>Ability of the child to coordinate respiration, phonation, resonation, and articulation to produce the sound system (phonology) of their language.</td>
<td></td>
</tr>
<tr>
<td>Domain</td>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Voice/Fluency</td>
<td>Areas related to the quality of a child’s speech including the rhythm, rate, pitch, loudness, and quality.</td>
<td></td>
</tr>
<tr>
<td>Oral Motor</td>
<td>Areas related to structural or functional factors that may affect communication and/or feeding including the articulators and oral and nasal cavities.</td>
<td></td>
</tr>
<tr>
<td><strong>Sensorimotor</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functions Underlying Movement</td>
<td>Automatic postural reactions (e.g., balance), muscle tone, ability to move body parts independently.</td>
<td></td>
</tr>
<tr>
<td>Gross Motor Activity</td>
<td>Areas related to large motor actions and positions including child’s preference for gross motor activity, quality of movement, level of independence, ability to assume and maintain developmental positions, effectiveness and efficiency of gross motor actions, coordinated use of two body sides,</td>
<td></td>
</tr>
</tbody>
</table>
and ability to move from one play activity to another.

**Arm and Hand Use**

Areas related to upper extremity skills and eye-hand coordination including child’s preference for fine motor activity, level of independence, effectiveness and efficiency of reach, hand and finger use, grasp, in-hand manipulation, and release.

**Motor Planning**

Child’s ability to figure out how to perform actions, use toys or equipment, and move through space under various conditions including modeling, verbal command, and self-initiated activity.

**Modulation of Sensation**

Areas related to child’s reactions to sensation including preference or avoidance of toys, space, and playmates; the match between reactions and the intensity of experiences; and the match between child’s general level of activity the demands of the situation.
<table>
<thead>
<tr>
<th>Domain</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Motor Contributions to Self Care</strong></td>
<td>Child’s performance on common daily activities requiring motor skill including ball games and other gross motor play (e.g., tricycle, swing); drawing, coloring, cutting with scissors and other pre-academic tasks; using toys and blocks; eating with utensils; and performing basic dressing tasks.</td>
</tr>
<tr>
<td><strong>Social-Emotional</strong></td>
<td></td>
</tr>
<tr>
<td>Emotional Expression</td>
<td>Communication of reactions, feelings, or intentions to others through facial patterns, muscle tension, body posture and position of extremities, movements, gestures, and words.</td>
</tr>
<tr>
<td>Emotional Style</td>
<td>The child’s typical affective response to different situations, including elements of temperament, such as approach or withdrawal to new stimuli and adaptability to change.</td>
</tr>
<tr>
<td>Domain</td>
<td>Definition</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>State Regulation</td>
<td>Ability to regulate physiological states of awareness (sleeping, crying, etc.) and control emotional reactions to both internal and external stimuli, including being able to inhibit impulsive actions and emotions and to self-calm.</td>
</tr>
<tr>
<td>Behavioral Regulation</td>
<td>The ability to control impulses, monitor one’s actions and interactions, and respond within the parameters of culturally accepted behavior, including compliance with adult requests, self-control over behaviors perceived as wrong, and use of social conventions.</td>
</tr>
<tr>
<td>Sense of Self</td>
<td>Understanding of self as a separate person capable of having an effect on his/her environment, including desire to accomplish goals and feel independent and competent.</td>
</tr>
<tr>
<td>Domain</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Social Relations</td>
<td>Ability to attend to social aspects of play, to read cues, to interpret and communicate social information, to get along with others and avoid negativity and conflict with others (including parents, strangers, siblings, and peers) within isolated, parallel, associative, cooperative or complementary roles in play interactions.</td>
</tr>
<tr>
<td>Emotional Themes of Play</td>
<td>Expression of inner feelings, worries, fears, and traumas through the actions of play, including through the dramatic representations of self or dolls.</td>
</tr>
</tbody>
</table>
Appendix B

Parental Rating Scale of Typical Behavior

Child’s ID Number: ____________________

Assessment Type: ____________________

Please circle the number that corresponds to the extent to which your child’s performance during the current evaluation represent the skills you have seen at home, in school, or in the community. You may also write a comment in the spaces provided below.

For Example:

To what extent did your child’s performance in language comprehension (ability to understand language) during the assessment represent the language comprehension skills you have seen at home, in school, or in the community?  EXAMPLE ONLY

1) not at all
2) somewhat
3) pretty closely
4) accurately

Begin Here: Gross Motor

To what extent did your child’s performance in big movement skills (walking, hopping, jumping etc) during the assessment represent the big movement skills you have seen at home, in school, or in the community?

1) not at all
2) somewhat
3) pretty closely
4) accurately

Comment:

Fine Motor
To what extent did your child’s performance in small movement skills (grasping a crayon, picking up small objects with his/her fingers, using scissors etc.) during the assessment represent the small movement skills you have seen at home, in school, or in the community?
1) not at all
2) somewhat
3) pretty closely
4) accurately

Comment:

Expressive Language
To what extent did your child’s performance in expressive language (talking and/or communicating with words) during the assessment represent the expressive language skills you have seen at home, in school, or in the community?
1) not at all
2) somewhat
3) pretty closely
4) accurately

Comment:
Receptive Language

To what extent did your child’s performance in language comprehension (understanding sounds, words, or gestures) during the assessment represent the language comprehension skills you have seen at home, in school, or in the community?

1) not at all
2) somewhat
3) pretty closely
4) accurately

Comment:

Behavior

To what extent was your child’s behavior (activity level, emotions) during the assessment represent the behavior you have seen at home, in school, or in the community?

1) not at all
2) somewhat
3) pretty closely
4) accurately

Comment:

Attention

To what extent was your child’s performance in attention span (the amount of attention given) during the assessment represent the attention span you have seen at home, in school, or in the community?

1) not at all
2) somewhat
Problem Solving

To what extent was your child’s performance in problem solving during the assessment represent the problem solving you have seen at home, in school, or in the community?

1) not at all
2) somewhat
3) pretty closely
4) accurately

Comment:

Is there anything that your child does at home that was not observed today during the evaluation that you would have liked the evaluator to observe?

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

This questionnaire was approved by the University of Denver's Institutional Review Board for the Protection of Human Subjects in Research on 6/10/08.
Appendix C

Test-Taking Behavior Questionnaire for Administrators

Team Member (Check all who are part of your team)

(1) Administrator of Bayley- III
(2) Clinical Psychologist or Social Worker
(3) Occupational Therapist/Physical Therapist
(4) School Psychologist
(5) Speech and Language Pathologist
(6) Teacher
(7) Other: ____________________

Child’s ID Number: _____________________
Child’s Date of Birth: ________________

Please circle the number below that corresponds to the child’s developmental behavior during the evaluation.

Cognitive Behavior

Attention to Tasks

Constantly off task; does not attend. 1
Typically off task; attends in one or two instances 2
Off task half the time. 3
Typically attends; attention wanders in one or two instances 4
Constantly attends. 5
**Adaptation to Change in Test Materials**

Consistently resists relinquishing materials and/or refuses to accept new materials.  
1

Typically resists relinquishing materials and/or refuses to accept new materials; makes one or transitions easily.  
2

Makes poor transitions half the time; makes good transitions half the time.  
3

Typically relinquishes materials and accepts new materials; one or two poor transitions.  
4

Consistently relinquishes materials and accepts new materials.  
5

**Interest in Test Materials and Stimuli**

No interest.  
1

Typically shows no interest.  
2

Shows initiative half the time.  
3

Typically shoes initiative; one or two instances of no initiative.  
4

Consistently shows initiative.  
5

**Language Behavior**

**Vocal Expressions**
Consistently inappropriate. 1

Typically inappropriate; one or two instances of appropriate vocal expressions. 2

Inappropriate half the time; appropriate half the time. 3

Typically appropriate; one or two instances of inappropriate vocal expressions. 4

Consistently appropriate. 5

**Useful gesture/pointing**

Consistently inappropriate. 1

Typically inappropriate; one or two instances of appropriate gesturing or pointing. 2

Inappropriate half the time; appropriate half the time. 3

Typically appropriate; one or two instances of inappropriate gestures or pointing. 4

Consistently appropriate. 5

**Understanding what others are saying**

Constantly misunderstands. 1

Typically misunderstands; one or two instances of understanding. 2

Misunderstands half the time. 3

Typically understands; misunderstands in one or two instances. 4
Constantly understands.  

**Social Emotional Behavior**

**Orientation to the Examiner**

Consistently avoids or resists, never responds.  

Typically avoids or resists; one or two instances of responsiveness.  

Avoids or resists half the time; responds half the time.  

Typically responds; one or two instances of avoidance or resistance.  

Consistently responds; never avoidant or resistant.  

**Hyperactivity**

Consistently hyperactive; fidgety and agitated in movement.  

Typically hyperactive; one or two instances of appropriate activity level.  

Hyperactive half the time; appropriate activity level half the time.  

Typically not hyperactive; one or two instances of hyperactivity.  

Consistently not hyperactive; never fidgety or agitated in movement.
Persistence in Attempting to Complete Tasks

- Constantly lacks persistence. 1
- Typically not persistence; one or two instances of persistence. 2
- Lacks persistence half the time. 3
- Typically persistent; lacks persistence in one or two instances. 4
- Constantly persistent. 5

Motor Behavior

Fine-Motor Movement Required by Tasks

- Consistently inappropriate. 1
- Typically inappropriate; one or two instances of appropriate fine-motor movement. 2
- Inappropriate half the time; appropriate half the time. 3
- Typically appropriate; one or two instances of inappropriate fine-motor movement. 4
- Consistently appropriate. 5

Gross-Motor Movement Required by Tasks

- Consistently inappropriate. 1
- Typically inappropriate; one or two instances of appropriate gross-motor movement. 2
- Inappropriate half the time; appropriate half the time. 3
Typically appropriate; one or two instances of inappropriate gross-motor movement.  
Consistently appropriate.

Hypersensitivity to Test Materials

Constantly hypersensitive: hypersensitivity disrupts testing.
Typically hypersensitive; returns to test activity in one or two instances.
Occasionally hypersensitive.
Typically reacts appropriately; hypersensitive in a few instances.
Constantly responds appropriately.

This questionnaire was approved by the University of Denver's Institutional Review Board for the Protection of Human Subjects in Research on 6/10/08.
Appendix D
Research Introduction Letter to Parents

Dear Parent:

My name is Keri Linas, and I am a doctoral student at the University of Denver in the College of Education. I am currently working on my dissertation project which seeks to determine what assessment approaches yield the most accurate information for children with and without a variety of special needs.

I would like to invite you and your child to participate in my project. If you choose to participate, you and your child would be asked to come to the Fisher Early Learning Center on a second time before or after the play evaluation. Your child would participate in a second developmental assessment lasting approximately 1-½ hours. If your child is between the ages of one and three years old, your child will be assessed with the Bayley Scales of Infant and Toddler Assessment. If your child is between the ages of three and one half and six, your child will be assessed with the Mullen Scales of Early Learning. These are the most widely used evaluation tools for professionals working with young children. Your child would interact with a variety of objects such as blocks, dolls, tea cups, and paper and pencil. The play assessment (which you have already scheduled) is called the Transdisciplinary Play Based Assessment. This is also a widely used evaluation. Your child will play with an adult facilitator while a team of professionals observe his/her behavior. During both evaluations, you will be in the room with your child.
You will also be asked to complete two questionnaires in addition to the paperwork you are asked to complete as a participant during the Transdisciplinary Play-Based Assessment Summer Institute. The questionnaires will take approximately 30 minutes to complete.

While your child will participate in activities that assess his/her strengths and weakness, no formal diagnostic testing for placement will take place, nor will anyone have access to any identifying information regarding your child except his/her first name and age. You would have the opportunity to receive information about your child’s development, whether you prefer one type of test to an other, whether your child feels more comfortable in one assessment or the other; and whether one assessment better captures your child’s skills. Your child will receive a book for his/her participation; you will receive a $5.00 gift card to Target, and you will be entered into a lottery for a $50 gift certificate to a local toy store.

Your participation is strictly voluntarily. If at any time during the research study, you wish to discontinue participation, you may do so with without any consequences. If you have any questions, please contact Keri Linas, MA or Dr. Toni Linder at 303-871-2474. Thank you for being a part of research that will help professionals in the education and psychological fields serve young children and their families. I appreciate your time and consideration.

Sincerely,

Keri Linas, MA
This letter was approved by the University of Denver's Institutional Review Board for the Protection of Human Subjects in Research on 6/10/08.
Appendix E

Parent-Child Informed Consent

Informed Consent Form

Participant Name: ______________________

Participant’s Date of Birth: ______________________

Date: ______________________

Name of Project:

Concurrent Validity of the Transdisciplinary Play Based Assessment-2 with the Bayley Scales of Infant and Toddler Development-III and the Mullen Scales of Early Learning.

Researcher’s Name: Keri Linas, MA.

You and your child are invited to participate in a study conducted as part of the Doctoral Dissertation of Keri Linas, MA through the Morgridge College of Education at the University of Denver. The study will be supervised by Dr. Toni Linder, Ed.D, Professor, Morgridge College of Education, University of Denver, (303) 871-2474 and Dr. Lavita Nadkarni, Professor, Graduate School of Professional Psychology, University of Denver, (303) 871-3877.

Purpose:

There are four purposes of the current study. First, the current study seeks to compare the ability of a standardized child development assessment and Transdisciplinary Play Based Assessment to provide accurate information about children’s development to help identify children in need of early intervention services. Second, the study seeks to examine how these assessments are connected to every day
behaviors like using a fork to feed oneself. Third, the current study seeks to understand whether there is a connection between performance on developmental evaluations and behavior during the evaluation. Finally, the current study seeks to determine whether certain types of evaluations can help the child demonstrate his or her most typical skills.

Description of the Study:

If you choose to participate in this study, you and your child’s participation will occur over a two-week period. You and your child would be asked to come to the Fisher Early Learning Center on two separate occasions. Your child would participate in two separate child development assessments lasting approximately one to two hours each. The use of standardized assessments would depend on your child’s age. If your child is between 1-36 months, your child would be assessed with the Bayley Scales of Infant and Toddler Assessment- Third Edition. If your child is over 36 months, your child would be assessed with the Mullen Scales of Early Learning. Your child would be assessed using one of these tools. These are the most widely used evaluation tools for professionals working with young children. Your child would interact with a variety of objects such as blocks, dolls, teacups, and paper and pencil. The other assessment is called the Transdisciplinary Play Based Assessment-Second Edition. Your child would play with an adult facilitator while a team of professionals observed his/her behavior. You will be present in the room during all assessments with your child.

During each assessment, the evaluator or team will complete test protocols, which help to identify your child’s strengths and needs. While your child will participate in activities that assess his/her strengths and needs, no formal diagnostic testing for placement will take place, nor will anyone have access to any identifying information.
regarding your child. In addition, to insure quality insurance, the evaluations will be recorded using a digital video recorder. You or your child’s name will not be associated with the recording.

Following the assessment you would be asked to complete two additional questionnaires beyond the requirements of the Summer Institute. One questionnaire asks the degree to which your child has obtained particular skills (e.g., talking back and forth with you) it is called the Vineland Scale of Adaptive Behaviors. The second questionnaire asks about your perceptions of the assessment (e.g., how accurately it captured your child’s skills, etc).

Participation in this project is strictly voluntary. Should you or your child experience discomfort during the assessment process you may discontinue your participation at any time. Refusal to participate or withdrawal from participation will involve no penalty or loss of benefits to which you are otherwise entitled.

Potential Risks:

The risks associated with this project are minimal. Assessments will take place during working hours and scheduling might be difficult. Your child may experience some anxiety interacting with an unknown adult. Developmental assessments can be stressful for some children. If, however, your child becomes upset during the evaluation, the procedure will be discontinued immediately.

Potential Benefits:

Upon completion of each assessment, your child will receive a book for his/her participation. At the completion of the study you will receive a $5.00 gift card to Target and you will be entered into a drawing for a $50 gift certificate to a local toy store. If you
choose to withdraw from the study for any reason, you will still be entered into the drawing for the gift certificate. The final drawing will occur when all of the assessments and corresponding forms have been received from all of the participations involved in the study.

You will also have the opportunity to receive two developmental evaluations, providing you with in depth information about your child’s development. You will learn which assessment process is more appropriate for your child. You will be able to use the information learned from the evaluations to communicate with his/her teachers for the upcoming school year.

Confidentiality:

Your responses will be identified by code number only and will be kept separate from information that could identify you. This is done to protect the confidentiality of your responses. Only the researchers will have access to your individual data and any reports generated. Results of this study will use only group averages and paraphrased wording. However, should any information contained in this study be the subject of a court order or lawful subpoena, the University of Denver might not be able to avoid compliance with the order or subpoena. Although no questions in the current study address it, we are required by law to tell you that if information is revealed concerning suicide, homicide, or child abuse and neglect, it is required by law that this be reported to the proper authorities.

If you have any concerns or complaints about how you were treated during the study, please contact Dr. Susan Sadler, Chair, Institutional Review Board for the Protection of Human Subjects, at 303-871-3454, or Sylk Sotto-Santiago, Office of
Sponsored Programs at 303-871-4052 or write to either at the University of Denver, Office of Sponsored Programs, 2199 S. University Blvd., Denver, CO 80208-2121.

You may keep this page for your records. Please sign below if you understand and agree to the above. If you do not understand any part of the above statement, please ask the researcher any questions you have.

I have read and understood the foregoing descriptions of the study called Concurrent validity of the transdisciplinary play based assessment-2. I have asked for and received a satisfactory explanation of any language that I did not fully understand. I agree to participate in this study, and I understand that I may withdraw my consent at any time. I have received a copy of this consent form.

Signature __________________ Date __________________

I would like a summary of the results of this study to be mailed to me at the following postal or e-mail address:

I agree to have my child video-taped for quality insurance

__________________________

Signature/ Date

This form was approved by the University of Denver's Institutional Review Board for the Protection of Human Subjects in Research on 6/10/08.
Appendix F

Child and Professional Demographics

Child Name: ___________________
Child Date of Birth: ________________
Child Sex: ___________________
Child Ethnicity: ___________________
Child Disability or Documented Delay : ______________
Family Approximate Income: ______________

How did you hear about the Summer Institute? (e.g., your preschool, early interventionist, doctor’s office, etc.) _________________

Professional Name: __________
Age:___________
Ethnicity: ___________
Years of Experience: __________
Discipline: __________

This questionnaire was approved by the University of Denver's Institutional Review Board for the Protection of Human Subjects in Research on 6/10/08.
Appendix G

Research Introduction Letter to Professionals

Dear Summer Play Institute Participant:

My name is Keri Linas, and I am a doctoral student at the University of Denver in the Morgridge College of Education. As a doctoral student, I am currently working on my dissertation project, which seeks to understand the best way for professionals like you working with young children to evaluate children’s development. This project is being supervised by Drs. Toni Linder, Ed.D and Lavita Nadkarni, Ph.D.

I am writing to invite you to participate in my project while you attend the Summer Play Institute. Your participation will help future professionals choose appropriate methods to understand children’s problem solving, language, speech, motor, and emotional development.

If you choose to participate during the Summer Play Institute, you will receive the standard Summer Play Institute curriculum. There would be minimal additional time demands. By participating you would agree to allow the assessment results to be used as data for the research project from each of the Transdisciplinary Play Based Assessments you complete to meet the Summer Play Institute curriculum requirements. You would also agree to complete a questionnaire about the child’s behavior during the Transdisciplinary Play Based Assessment. This questionnaire will take approximately ten minutes. At the end of the Summer Play Institute, to honor your participation in the research study, you will receive a $5.00 gift card to Target.

Your participation, however, is not a mandatory component to the Summer Play Institute. It is strictly voluntary and if at any time during the Summer Play Institute, you
wish to discontinue participation, you may do so without any consequences. If you have any questions, you may contact Keri Linas or Dr. Toni Linder at 303-871-2474.

Thank you for being a part of research that will help professionals in the education and psychological fields learn how to better serve young children and their families. I appreciate your time and consideration.

Sincerely,

Keri Linas, MA.

This letter was approved by the University of Denver's Institutional Review Board for the Protection of Human Subjects in Research on 6/10/08.
Appendix H

Professional Informed Consent

Informed Consent Form

Name of Project:
Concurrent Validity of the Transdisciplinary Play Based Assessment-2 with the Bayley Scales of Infant and Toddler Development-III and the Mullen Scales of Early Learning.

Researcher’s Name: Keri Linas, MA.

You are invited to participate in a study conducted as part of the Doctoral Dissertation of Keri Linas, MA through the University of Denver. The study will be supervised by Dr. Toni Linder, Ed.D, Professor, Morgridge College of Education, University of Denver, (303) 871-2474 and Dr. Lavita Nadkarni, Professor, Graduate School of Professional Psychology, University of Denver, (303) 871-3877.

Purpose:

There are four purpose of the current study. First, the current study seeks to compare the ability of a standardized child development assessment and Transdisciplinary Play Based Assessment to provide accurate information about children’s development to help identify children in need of early intervention services. Second, the study seeks to examine how these assessments are connected to every day behaviors like using a fork to feed oneself. Third, the current study seeks to understand whether there is a connection between performance on developmental evaluations and behavior during the evaluation. Finally, the current study seeks to determine whether certain types of evaluations can help the child demonstrate his or her most typical skills.
Description of the Study:

If you choose to participate in this study, you will complete the formal curriculum of the Transdisciplinary Play Based Assessment Summer Institute, which includes completing assessments with children and their caregivers and scoring the assessments. You will be asked to submit for the purpose of research your scored Transdisciplinary Play Based Assessment results without the name of the child. Following each assessment, you will also be asked to work in your assessment team to answer a questionnaire about the child’s behavior during the evaluation (e.g., How accurately it captured the child’s skills, etc).

In addition, to insure quality insurance, the evaluations will be recorded using a digital video recorder. Your name will not be associated with the recording.

Your participation in this research study is not a mandatory component to the Summer Play Institute. It is strictly voluntary and if at any time during the Summer Play Institute, you wish to discontinue participation in the study, you may do so with without penalty or loss of benefits to which you are otherwise entitled.

Potential Risks:

The risks associated with this project are minimal. It may take and additional 10 to 15 minutes to complete the child behavior questionnaire.

Potential Benefits:

Upon completion of the study, to honor your participation you will receive a $5.00 gift card to Target. If you choose to withdraw from the study for any reason, you will receive the gift card. You will also have the opportunity to receive the results of the
current study which may help inform future decisions on the type of assessment you use with young children.

**Confidentiality:**

Your responses will be identified by code number only and will be kept separate from information that could identify you. This is done to protect the confidentiality of your responses. Only the researchers will have access to your individual data. Any reports generated as a result of this study will use only group averages and paraphrased wording. However, should any information contained in this study be the subject of a court order or lawful subpoena, the University of Denver might not be able to avoid compliance with the order or subpoena. Although no questions in this project address it, we are required by law to tell you that if information is revealed concerning suicide, homicide, or child abuse and neglect, it is required by law that this be reported to the proper authorities.

If you have any concerns or complaints about how you were treated during the study, please contact Dr. Susan Sadler, Chair, Institutional Review Board for the Protection of Human Subjects, at 303-871-3454, or Sylk Sotto-Santiago, Office of Sponsored Programs at 303-871-4052 or write to either at the University of Denver, Office of Sponsored Programs, 2199 S. University Blvd., Denver, CO 80208-2121.

You may keep this page for your records. Please sign below if you understand and agree to the above. If you do not understand any part of the above statement, please ask the researcher any questions you have.

I have read and understood the foregoing descriptions of the study called Concurrent Validity of the Transdisciplinary Play Based Assessment-2. I have asked for
and received a satisfactory explanation of any language that I did not fully understand. I agree to participate in this study, and I understand that I may withdraw my consent at any time. I have received a copy of this consent form.

Signature _____________________ Date _________________

I agree to be video-taped for quality insurance ____________________

Signature/ Date

I would like a summary of the results of this study to be mailed to me at the following postal or e-mail address:

This consent was approved by the University of Denver's Institutional Review Board for the Protection of Human Subjects in Research on 6/10/08.
Appendix I
Research Assistant Protocol

August 4th

1) Introduce self to the professionals in the morning and your role. Briefly explain
   rational for the study (See Below). Request volunteers who have not already
   volunteered. Provide details of their role in the study. Answer any questions that
   they have, if you can’t answer, don’t make it up. Let them know you will get back
   to them. Call Keri with questions.

2) Set up Play Room for Bayley-III or MSEL. (See below).

3) Begin conducting the Bayley-III/MSEL Assessment. (See below).
   a. Introduction to Parent.
   b. Administer Bayley-III/MSEL.
   c. Give parent measures (2).
   d. Give book to child.
   e. Remind parent when they are returning for TPBA-2. Provide informal
      feedback.
   f. Collect parent measures.
   g. You complete Test-Taking Behavior Rating Form (See Below)
   h. You score Bayley-III/MSEL, note if child needs referral, look up
      appropriate number on provided PDF form of Child Find. Plan to give
      this information to parent the end of the Institute.
4) Explain TPBA-2 scoring procedures to the professionals, 10 minutes before class breaks for lunch.

5) Place scored measures upstairs.

August 5th

1) Continue with Bayley-III/MSEL Assessment (repeat above).

2) Check in with the professionals to see if they have questions about their responsibilities for August 6th-8th.

3) Placed scored measures upstairs.

August 6th

1) Introduce self to the parents who have not yet come to Fisher. There will also be parents who come to the institute who have not volunteered for study participation either because they do not want to participate or because they forgot to send their form. Please distribute the study recruitment letter to parents who have not signed up for the study (take no as answer). (See below).

2) Collect consent forms from parents who are new to the study- schedule their Bayley-III/MSEL for the following week.

3) Give Child Find Referral phone numbers and contact info to parents of children who were already assessed with the Bayley-III/MSEL who need further evaluation.
4) Professionals complete TPBA-2. (See protocol).
   a. Professional score TPBA-2.
   b. Give professionals, as a team, Test-Taking Behavior Rating Form, to be completed as a “team.”
   c. Give parent measures.

5) Give gift card to parents whose child has completed the Bayley-III/MSEL and the TPBA-2.

6) Place scored measures upstairs.

August 7th
1) Repeat steps 1-6.

August 8th
1) Repeat steps 1-6.
2) Give gift cards to professionals who have participated.
3) For children who have completed both assessments, mark their name off so that we can enter them into the raffle.

August 11th – 12th
1) Administer and score Bayley-III/MSEL, mark child’s name as complete.
2) Give gift card and book.
3) Score Vineland-II.

1. Introductions August 4th.
   a. To Professionals
Hello, my name is xxx, I am a [masters] student in the Child Family School Psychology Program here at DU. I am working with a doctoral student, Keri Linas, who is currently working towards her dissertation. She is interested in the validity of developmental evaluations for young children. Specifically whether play assessments and standardized assessments can get the same information, which type of evaluation succeeds at eliciting the child’s most typical range of skills, and how test behavior relates to children’s test scores.

*Please Read if more volunteers needed:* We are still looking for volunteers to participate in the study. You are not obligated to participate and if you do not participate you will still receive all of the benefits from the Institute. I am going to take five minutes to explain your responsibilities and then individuals can approach me if they would like to volunteer.

*(Brief Explanation of the Study) Please Read:* You (i.e., professionals at the Institute) will have two responsibilities. 1) Score the TPBA-2 of the child you assess as a team and provide me with your results. 2) You will be asked to complete a behavior-rating tool of the child’s behavior during the assessment with your team members. Only those team members that are participating in the study will complete the behavior-rating tool. In honor of your participation, you will be given a gift card to Target.

To all research participants, I will meet with you today 10 minutes before for the lunch break and explain the scoring procedures.

2. Set up Room for Bayley-III/MSEL.
   a. A child size table and two chairs.
   b. Place a “Do not disturb” sign on playroom door.
c. Place child chair with back of chair against the cabinets along the wall.

d. Make sure table is close to chair.

e. Avoid wearing bright or stimulating jewelry, clothes, or accessories.

3. Introduction To Parents (Bayley-III/MSEL)

Hello, my name is xxx, I am a (masters) student in the Child Family School Psychology Program. Thank you for agreeing to participate in this study. Today will take about an hour. If you have questions, please interrupt me at any time. As you may already know, this study is a two (day or week) process. Each session will last approximately one hour. You will return to this location (on X or next week on X). You and your child’s participation are strictly voluntarily and you may discontinue at any time. This information will be kept confidential. This information will not be used for formal diagnostic labeling or intervention planning. I will give you some informal feedback. If I believe it might be helpful for your child to be referred for further evaluation, I will provide you with that contact information at your next visit.

Today I am going to work with your child in our playroom. I will administer the Bayley Scales of Infant and Toddler Development. Have you ever heard of it, or seen it? It is an assessment tool that allows professionals to obtain information about your child’s development. Your child will interact with objects such as blocks, dolls, puzzle shapes, and cups. You are more than welcome to sit in the room during the assessment. Depending on your child, he/she may even wish to sit in your lap. The only thing that I ask is that you try to refrain from helping your child in his/her
interactions with the objects. I want to see what he/she is able to do before adults help him/her. After the assessment today, I will ask you to complete a brief questionnaire about you’re your child’s today. This will take approximately 15 minutes. I will also ask you to complete a second questionnaire about your child’s general behavior and skills. This will take approximately 25 minutes. Your child will receive a book today for his/her participation.

Next week (or on X day of this week), you and your child will return to the Fisher Early Learning Center. Your child will participate in the second assessment. You will receive more details about that when you return. At the end of the study, you will receive a $5.00 gift card to Target and be entered into a drawing for a $50 gift certificate to a local toy store.

4. Obtain Consent Form

Ask parent to sign consent form if she/he has not already provided it to you. Ask child if she/he would like to see the playroom and see some toys you have for her/him.

5. Administer Bayley-III/MSEL
   a. Parent can be in the room with child, ask child to ‘assent’ to participate.
   b. If you need to provide stickers, to help child during the assessment, that is ok, take note what behavior strategies you had to use.
   c. If child does not need parental support, parent can begin completing the social emotional scale of the Bayley that require parent completion. They can also begin the Vineland-II (parent measure # 1). If child needs
parents support, wait until after Bayley-III is completed. MSEL does not have parent measure.

d. After you are done giving Bayley-III, ask parent to finish the Bayley-III parent measures- to complete the Vineland-II (if she/he has not already completed it) and give parent the “Typical Behavior Rating Form” (parent measure # 2). This measure must be given after the assessment.
   
i. Read the Instructions to the Vineland-II, prior to giving.
   
   ii. Read the Instructions to the Bayley-III Caregiver prior to giving.

   e. Give child a book and sticker for participation.

   f. Ask parents if there are immediate questions. Explain to parent that you will let the parent know when they return that if you think the child needs a follow-up evaluation, you will provide the details at the next visit.

   g. Remind parents that this assessment is not being used for diagnostic/intervention purposes.

   h. Score Bayley-III/MSEL according to the manual; only use child’s number.

   i. You complete the Child Test Taking Behavior.


   k. Place upstairs in cabinet.

6. Introduction to Parents TPBA-2

   (For parents/children who have not yet completed the Bayley-III/MSEL and have not sent in consent forms either because they forgot or they do not want to participate)
Hello, my name is xxx I am a (masters) student in the Child Family School Psychology Program here at DU. I am working with a doctoral student, Keri Linas, who is currently working towards her dissertation. She is interested in the validity of developmental evaluations for young children. Specifically whether play assessments and standardized assessments can get the same information, which type of evaluation succeeds at eliciting the child’s most typical range of skills, and how test behavior relates to children’s test scores.

We are still looking for parent/child volunteers to participate in the study. Here is a letter that explains what you and your child would do. Please take a minute to look it over. You are not obligated to participate and if you do not participate you will still receive all of the benefits from the Institute. Let me know if this is something that might interest you or there are questions I can answer (If parent is interested, ask parent to sign consent form).

To all parents (those who just gave consent and those for whom you have consent). Thank you for agreeing to participate in this study. Today will take about an hour. If you have questions, please interrupt me at any time. As you may already know, this study is a two-week process. Each session will last approximately one hour. You and your child’s participation are strictly voluntarily and you may discontinue at any time. This information will be kept confidential. This information will not be used for formal diagnostic labeling or intervention planning.

Today your child will play with a professional while a team of child development professionals observes his/her behavior. This process is called a Transdisciplinary Play Based Assessment. Have you ever heard of it, or seen it? It is an assessment tool that
allows professionals to obtain information about your child’s development. Your child will interact with toys such as dolls, puzzle shapes, dress up clothes, trains etc. You are more than welcome to sit in the room during the assessment. The only request that I have is that you try to refrain from helping your child in his/her interactions with the objects. I want to see what he/she is able to do before adults that he/she knows, help him/her.

After the assessment today, I will ask you to complete a brief questionnaire(s) about your child’s experience today (as well as his/her general development). It will take approximately twenty minutes. Your child will receive a book today for his/her participation.

Next week you will return to the Fisher Early Learning Center. Your child will participate in the second assessment. You will receive more details when you return. At the end of the study, you will receive a $5.00 gift card to Target and be entered into a drawing for a $50.00 gift certificate to a local toy store).

7. Obtain Consent Form

Ask parent to sign consent form if she/he has not already provided it to you. Ask child if she/he would like to see the playroom and see some toys you have for her/him.

8. Teams complete TPBA-2

   a. Team scores TPBA-2.
   
   b. As a team, complete the Test-Taking Behavior Rating tool.
   
   c. Parent completes the Typical Behavior tool.
d. Parent completes the Vineland-II (if parent has not yet completed this tool).

e. You give child a book.

f. You give parents who have completed both Bayley-III/MSEL and TPBA-2 a gift card. Provide referral to parent if needed.

g. Collect the scored TPBA-2 and measures. Make sure they are clipped together by child. Place upstairs.