

University of Montana

ScholarWorks at University of Montana

Graduate Student Theses, Dissertations, &
Professional Papers

Graduate School

1989

Measurement of interobserver reliability and concurrent validity for the Play Assessment Scale

Kari Lee Altenhofen
The University of Montana

Follow this and additional works at: <https://scholarworks.umt.edu/etd>

Let us know how access to this document benefits you.

Recommended Citation

Altenhofen, Kari Lee, "Measurement of interobserver reliability and concurrent validity for the Play Assessment Scale" (1989). *Graduate Student Theses, Dissertations, & Professional Papers*. 5708.
<https://scholarworks.umt.edu/etd/5708>

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.

MEASUREMENT OF INTEROBSERVER RELIABILITY
AND CONCURRENT VALIDITY FOR
THE PLAY ASSESSMENT SCALE

By
Kari Lee Altenhofen
Thesis presented in fulfillment of the
requirements for the degree of
Master of Arts
University of Montana
1989

Approved by

Barbara Bacon

Chairman, Board of Examiners

W. C. Murray

Dean, Graduate School

December 18, 1989

Date

UMI Number: EP41175

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI EP41175

Published by ProQuest LLC (2014). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

1-8-90

Altenhofen, Kari, L., MA., March 1990

Communication
Sciences and Disorders

Measurement of Interobserver Reliability and Concurrent
Validity for the Play Assessment Scale

Director: Barbara Bain, Ph.D., BB

This study addressed a concern in the area of assessment of young infants and preschool children. This concern relates to a lack of psychometric data, such as, concurrent validity and reliability, to support the use of many assessment procedures used with this population. The purpose of this study was to examine the relationship between "play age" obtained on the Play Assessment Scale and "age equivalent" obtained on the Battelle Developmental Inventory Cognitive and Communication Domains. These two assessment procedures were administered to 18 children, 38% of whom had handicapping conditions and 61% of whom were normally developing. The age range of these subjects was between 4 and 53 months. The mean age of participants was 25.5 months. Two clinicians administered these assessment procedures. Administration of the two assessment procedures were counterbalanced for order of presentation. Assessment procedures were videotaped and independently scored by examiners to determine interobserver reliability. A Pearson Product Moment Correlation Coefficient was used to determine the relationship between the scores obtained on the two assessment procedures.

Results of this study indicated high interrater reliability (.96) for both Cognitive and Communication Domains of the Battelle Developmental Inventory and moderately high interobserver reliability (.77) for the Play Assessment Scale. The correlation coefficients between the Play Assessment Scale and Cognitive and Communication Domains of the Battelle Developmental Inventory were high (.9279 and .9369) and strongly support the use of the Play Assessment Scale. The importance of the concurrent validity results was discussed in terms of infant and preschool assessment. Suggestions for increasing interobserver reliability were provided.

Acknowledgements

I wish to thank Don Goldberg and Barb Bain for their guidance, assistance and dedication as I developed this thesis. I would also like to thank Sarah Mulligan, Kathy Miller and Sandra Morris for their support and the many hours of testing, videotaping and scoring they did. I am much in debt to my friend Jim for his assistance in "computer land". Last, but not least, I would like to offer my thanks to the parents and children who participated in this study.

Table of Contents

	Page
Abstract.....	ii
Acknowledgments.....	iii
List of Tables	v
Chapter 1 Introduction.....	1
Chapter 2 Methods	37
Chapter 3 Results	42
Chapter 4 Discussion	50
Bibliography	74

List of Tables and Figures

Table	Page
1. Individual Subject Characteristics.....	38
2. Correlation coefficients between the "play ages" of the <u>Play Assessment Scale</u> and the "age equivalencies" of the <u>Battelle Developmental Inventory</u> Cognitive and Communication Domains.....	45
3. Interobserver reliability measures for the <u>Battelle Developmental Inventory</u> Cognitive and Communication Domains.....	48
4. Interobserver Reliability Measures for the <u>Play Assessment Scale</u>	49

Figure

1. Scattergram of <u>PAS</u> "play age" scores as a function of <u>BDI</u> "age equivalent" scores.....	46
---	----

Chapter I

INTRODUCTION

Infant and Preschool Assessment

An increasing emphasis on early education of young children with handicaps is occurring, particularly with the advent of P.L. 99-457 and its implications for the education of children from birth to three. In a recent article, Sheenan (1989) discussed the implication of P.L. 99-457 for assessment with young children. More educators and clinicians will be responsible for assessing infants and toddlers as this law mandates services for this population. Sheenan (1989) stated, "yet another quantitative implication of P.L. 99-457 is that as services for preschoolers with handicaps increase, legal challenges to the validity of early childhood measures are also likely to increase." Professionals concerned with the assessment of this population will be interested in assessments which fulfil several requirements of P.L. 99-457; those being first to determine eligibility for programs and

then to determine goals and objectives for the Individual Family Service Plan (IFSP). Early interventionists are faced with the challenge of finding assessment procedures which will meet the needs of these young children. Many authors have discussed the purposes, concerns, and possible solutions to problems of assessment procedures used with young children with handicaps (Bailey and Worley, 1984; Bricker, 1982; Brookes-Gunn and Lewis, 1981; Garwood and Fewell, 1983; Hanson, 1984; Neisworth and Bagnato, 1988).

Johnson (1982) states that interventionists need "accurate, reliable, and useful ways to describe the developmental status of these youngsters." She states that assessment procedures are needed for the following phases of service: "identifying those in need of intervention, determining goals for intervention, charting developmental progress, and evaluating the effectiveness of intervention." Added to this list are several other reasons for conducting assessments: admission into a program, identification of strategies for instruction, to determine if change in program placement is warranted, (Fewell, in Garwood and Fewell, 1983);

and "to provide information regarding strengths and weaknesses" for intervention planning (Bricker and Littman, 1982, cited in Hanson 1984).

Several types of assessment procedures used in early intervention exist. Norm-referenced assessments compare a child's performance with the performance of normal children the same age (Fewell, 1983; Bailey and Worley, 1984; Neisworth and Bagnato, 1988). These tests are typically used for identification and placement considerations (Fewell, 1983); describing child's development compared to normative data, placement into a diagnostic category, or predicting the child's development (Neisworth and Bagnato, 1988).

Curriculum-based assessments are "a form of criterion-referenced evaluation" which "traces a child's achievement along a continuum of objectives" (Neisworth and Bagnato, 1988). A child's performance then is basically compared to his or her past performance to monitor progress. Purposes of this type of assessment include: identification of treatment goals, determination of strengths and weaknesses, charting progress, use by interdisciplinary team, and measurement of program

efficacy (Neisworth and Bagnato 1988).

Both norm-referenced and curriculum-based assessments would be considered forms of direct testing (Bailey and Worley, 1984). Indirect methods are becoming increasingly popular in the assessment of young children. Neisworth and Bagnato, (1988) describe a number of such measures including: process, judgement-based, ecological, interactive, and systematic observation assessment. These assessment procedures allow the examiner or observer to collect data about a child in regular, routine situations which are generally less structured than direct testing.

Bailey and Worley (1984) list some of the advantages and disadvantages of both direct testing measures and observational measures. The advantages of direct testing measures are, they allow "meaningful comparisons of children", are "necessary for diagnostic needs", and "facilitate the transfer of information". On the other hand, they usually contain "no adaptations for children with sensory or motor impairments", "lack of validated measures for educational planning", and "skills sampled are limited to those included

in the test". Observational procedures have the advantages of measuring "what children do in the real world settings", are "sensitive to changes over time", and "can be done during regular classroom activities". However, these procedures are "time-consuming", "require a certain amount of skill to design a good observation system", and "lack guidelines to interpret data gathered".

The field of early childhood intervention has grown rapidly during the last decade and with that growth comes the recognition that research is needed which will examine various aspects of assessment of infants and preschoolers with special needs (Brookes-Gunn and Lewis, 1981; Sheehan and Gallagher, 1984; and Johnson, 1982). Some concerns relating to existing assessment procedures used in early intervention are described below. The characteristics of multiply handicapped or special needs children may not be adequately described by the available test instruments. Often these assessment tools are the products of tests used for older or nonhandicapped populations (Johnson, 1982; Neisworth and Bagnato, 1988). Items on such tests are not appropriate for children with

significant motor and sensory deficits (Johnson, 1982; Garwood and Fewell, 1983). Johnson (cited 1982) states "norm-referenced tests used to measure program effectiveness show little or no progress for severely impaired because these children have sensory or motor problems that prevent appropriate responses to the demands of the test". Brookes-Gunn and Lewis (1981) report that some of the most popular assessment procedures usually represent the child's performance by "a single age equivalent score or an intelligence quotient, focus on developmental milestones, concentrate on one or two areas of development to the exclusion of others", and ignore the interaction of developmental domains".

Problems associated with assessment are not confined to norm-referenced, standardized procedures. Johnson (1982) lists these problems related to criterion-referenced tests; many of the "items are taken directly from the standardized tests they were designed to replace"; they often provide age levels of mastery, also taken from standardized tests; and these tests are less reliable because the items are taken from a number

of tests, with different standardization samples. Another major concern with current assessment procedures is the lack of various psychometric properties, including reliability, concurrent validity, predictive validity, and normative samples (Brookes-Gunn and Lewis, 1981; Sheenan, 1982). Bracken (1987) advocates that a critical look at preschool assessment procedures is warranted. He states, "concomitant with the increase in preschool assessments, there needs to be increased professional attention paid to the quality of instruments used in preschool assessment".

Problems inherent in assessing infants and young preschoolers under the age of three or four are associated not only with the available tests, but also with the characteristics associated with this population. For example, if the purpose of conducting an assessment of a child is diagnostic then the question of predictive validity may arise, particularly when assessing young children with multiple handicaps and/or children with sensory or neurological problems. These children may not respond to the traditional test situations or tasks requiring motor, language or

perceptual responses. Modifications of test items may provide the examiner with a clearer example of the child's abilities, but these modifications will most likely invalidate the test results (McLean and Snyder-McLean, 1978, cited in Bailey and Worley, 1984). Difficulties may exist with motivation, not only related to the test materials, but the attention span of the child (Bailey and Worley, 1984). As young children are often uncomfortable with strangers, establishing rapport may be difficult (Bailey and Worley, 1984). In order to make the child feel secure and conduct the evaluation, the presence of a parent may be required which leads to the possibility of "coaching" by the parent. Young children may often possess their own agenda for how to interact with the test materials or compliance may be an issue. These factors may result in inconsistencies of performance. Bailey and Worley (1984) refer to this as the "variable performance patterns" often exhibited by young children with handicaps. Factors contributing to variable performances are that children may be receiving "medication for seizures, hyperactivity, or illness," may exhibit unnoticed mild seizures during testing, may "have less

endurance than the average child, and may "exhibit rapid fluctuations in level of alertness". The lack of demonstration of skills or abilities may affect the reliability of scores or diagnosis.

In considering these problems with infant and preschool assessment, some writers (Brookes-Gunn and Lewis, 1981; Johnson, 1982; Fewell, 1983; Bailey and Worley, 1984; Neisworth and Bagnato, 1988) have proposed some solutions. Brookes-Gunn and Lewis (1981) propose that many developmental domains be emphasized and that interactions between domains be examined. Johnson (1982) advocates the development of procedures which provide modifications or adaptations for children with all types of impairment. She also suggests researchers conduct research studies examining the reliability and validity of existing measures. Fewell (1983) presented examples of new assessment measures and new directions in infant assessment. These new procedures offer more detailed information regarding child development, for example information can be gained regarding perceptual-cognitive performance through changes in heart rate or habituation

responses to taste, smell, light, sound and movement stimuli (Kagan, Kearsley and Zelano, 1978, cited in Johnson, 1982 and Fewell, 1983). Fewell (1983) also suggests the use of observational procedures that can be performed before or during the formal test situation. These informal observations can provide information regarding selection of formal tests, intervention planning and "particular strategies and techniques" to be used in assessment and in planning instructional programs. One possible area of observation proposed by Fewell was play. Observations of play and toy preference may increase motivation and compliance, as examiners will be aware of possible reinforcers and age-appropriate materials. Fewell suggests those conducting assessments begin with informal procedures before formal testing occurs. Bailey and Worley (1984) suggest using a combination of standardized, observational, and interview procedures for a comprehensive assessment. They further stress the importance of multiple assessments done over several days, the use of interdisciplinary assessments, and the importance of involving the parent in the assessment process.

Neisworth and Bagnato (1988) have identified a need to link developmental assessment with a number of other child related dimensions in what they term "multidimensional assessment". This refers to "a comprehensive and integrated evaluation approach that employs multiple measures, derives data from multiple sources, surveys multiple domains and fulfills multiple purposes." Using more than "one form of assessment can serve as a 'check' to challenge the accuracy of more traditional (but often misleading) forms of assessment." Traditionally, the areas of fine and gross motor, socio-emotional, self-care, communication and cognition are used to determine a child's status and to develop an intervention program. In a "multidomain approach" to assessment, not only are these traditional domains assessed, but dimensions such as, "mastery motivation", "social competence", "temperament", "self regulation", "attentional", "emotional expression", "early coping behavior" and "play" are also included. The current research study will focus on one such alternative assessment procedure, (Play Assessment Scale, Fewell, 1986).

Fewell (1986) has developed a preliminary procedure, designed to assess children's play in an unobstructed, observational format, called the Play Assessment Scale (PAS). The importance of child's play can not be overlooked. It spans across the domains of personal-social, problem-solving, motor, and communication. In addition, it can provide information regarding attention, child preference of materials, and level of prompting a child needs to demonstrate a particular play behavior, (based on spontaneous behavior in unstructured situations). The observation of play behaviors varies from other standard assessment procedures in that it allows the observer to follow the child's line of interest rather than have the examiner be in complete control of the assessment procedure.

The PAS results in an assigned play age, but there is no information on the normative sample from which this play age was derived. As of yet, the author of the PAS has not offered information on the validity or the reliability of this assessment procedure. As was discussed earlier, this is a pervasive problem with assessment procedures used in early

childhood education.

The PAS has the advantages associated with observational measures in that it is unobtrusive and makes use of routine play situations. Lack of attention and motivation are less likely to interfere with the results, increasing the likelihood of true or valid scores. The information gathered from the PAS would be useful in designing intervention programs, a characteristic which many formal tests do not offer. However, the PAS may also have the disadvantages associated with unstandardized, unstructured procedures (ie., questionable validity and reliability, problems with interpretation of the results, observations being dependent on the skills of the examiner). Given these advantages and disadvantages, the assessment of the reliability and validity of the PAS is an important consideration. As suggested by Rentz (1977, cited in Sheenan, 1983), "rather than develop another screening instrument, a preferable strategy would be to conduct validity studies on some of the more promising popular instruments". The purpose of this study, then, is to determine the concurrent validity between the PAS results and

results obtained using a relatively recent but, increasingly accepted assessment procedure, the Battelle Developmental Inventory (BDI) (Newborg, Stock, Wnek, Guidubaldi & Svinicki, 1984). The BDI includes five domains including social-emotional, personal, motor, communication and cognitive for assessment. It was not within the scope of this study to compare the results obtained on all five domains or the total BDI with results obtained on the PAS. The domains of communication and cognitive development were chosen on the basis that these are considered to be closely linked to play development. The nature of the relationship between play and cognition and play and language and the rationale for selecting these domains will be discussed later.

Rational for using the Battelle Developmental Inventory

An ideal assessment procedure would be one that not only could be used as a diagnostic tool, but also to determine the effects of an educational program on the progress of groups of children and develop individual program goals. This type

of assessment procedure has been cited as the "pre-eminent diagnostic/prescriptive approach in applied settings for young exceptional children" (Fewell and Sandall, 1983, cited in Neisworth and Bagnato, 1988). Optimally, an assessment procedure assists in the planning of a curriculum and the adaptation of teaching activities for children with delays in one or more areas of development. Thus, children's strengths and weaknesses across a variety of domains could be taken into account for the planning and implementation of play and learning activities, as well as daily care routines (eg., feeding, dressing, personal care, toileting). In addition, such measures could be used to group children with handicaps with developmentally appropriate peers for small group activities. One such developmental assessment tool is the Battelle Developmental Inventory.

Battelle Developmental Inventory

The Battelle Developmental Inventory (BDI) is a recently developed, well standardized assessment battery which is an example of a "norm-referenced diagnostic measure that also

integrates adaptive and curriculum-referenced features into its structure" (Neisworth and Bagnato, in press). These authors report the BDI can be used for both identification and intervention of children with handicaps, as well as a means to evaluate the effects of an educational program on the progress of children with handicaps. The BDI evaluates a large number of critical developmental skills across five domains and 22 subdomains. The five domains include: personal-social, adaptive, motor, communication, and cognition. The BDI is somewhat unique in that it includes specific adaptations (both stimulus and response) for children with sensorimotor impairments or other developmental disorders. It gives percentile ranks, Z-scores, T-scores, developmental quotients, age equivalents, and normal curve equivalents. In addition, the items are congruent with the developmental and behavioral goals and tasks of many popular infant and preschool curriculum materials. The "normative and technical data on the BDI strongly support its use" (Neisworth and Bagnato, 1988). The BDI reports comprehensive information regarding standardization. It was standardized on 800 infants,

preschoolers, and early school age children across the birth to 8 year age range. The standardization sample was stratified across geographic areas, ages, races, and sexes. The manual also reports sufficient statistical information regarding reliability (test-retest and interrater) and validity (content, construct, and criterion related) to support its stability for field and clinical use.

The authors report high and significant correlations for concurrent validity between the BDI and the Vineland Social Maturity Scale (Doll, 1965) and the Developmental Activities Screening Inventory (Dubose and Langley, 1977). They report moderate and positive correlations with the Stanford-Binet Intelligence Scale (Terman, and Merrill, 1960). Correlations with the Peabody Picture Vocabulary Test (Dunn and Dunn, 1981) were high, particularly with the expressive and receptive domains of the BDI. Correlations with the Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974) were lower, but should be interpreted cautiously as these were obtained on a small sample.

McClellan, McCormick, Bruder and Burdick (1987) reported high

concurrent validity between the BDI and the Bayley Scales of Infant Development and the Vineland Scales of Adaptive Behavior.

"The BDI is an excellent example of a multidimensional assessment battery which blends norm- and criterion-referenced features to link assessment and intervention. Its inclusion of adaptive evaluation strategies ensures the collection of more accurate and functional diagnostic and instructional data" (Neisworth and Bagnato, 1986). Harrington (1985) states the BDI meets not only the psychometric standards for educational and psychological testing (1974), but it also meets the letter and spirit of P.L. 94-142. Fewell, (in Mott et al., 1986) also recognizes the BDI as the best assessment instrument to be used in evaluating the outcome measures for early childhood special education programs. She listed nine practical aspects and applications which make it appealing including:

- a recently standardized screening test and comprehensive assessment across five domains,
- appropriateness for developing IEP's,
- use with individuals and groups,
- can be administered by program staff,

- easy to follow manual,
- adaptations for the handicapped,
- scoring system of 0, 1, and 2 which permits assessment of change in moderately and severely handicapped,
- provides a test profile across all domains.

Play Assessment Scale

The Play Assessment Scale (Fewell, 1986) was developed as a means for assessing children's play behaviors covering the age range from birth to three. The PAS provides the examiner with a list of suggested toys needed to observe a range of behaviors from sensorimotor exploration to symbolic play. The intent is to first observe those behaviors on an observational level so the child demonstrates those skills which are spontaneous and therefore determined to be within the child's play repertoire. The examiner can then determine what the child knows about his environment independent of adult interference. If a given behavior is observed spontaneously without any adult directed cues or prompts the child is given credit for that behavior. If a specific behavior is not noted to occur spontaneously, the examiner then begins to prompt the child to perform that behavior first

at a verbal level, then the behavior can be modeled and lastly the examiner can both model and verbally prompt the behavior. A child is given credit for the behavior only if it occurs spontaneously. The behaviors are then totaled and based on the total number of spontaneous behaviors observed, the child is assigned a play age. The advantage of coding the behaviors as occurring following a verbal, model, or verbal+model prompt is for the purposes of designing intervention strategies. The examiner will then know at what prompt level the child was able to demonstrate the behavior.

The items included in the PAS are the result of an extensive examination of research on play and seem to be influenced by the Piagetian theory of children's cognitive development as well as the typical developmental theory in which the development of "atypical" children is compared to the development of "normal children". Items on the PAS are divided into three major divisions. These being exploratory, representational, and symbolization. The earliest play behaviors observed in children are exploratory in nature meaning the child's actions are often reflexive, for example,

mouth, banging, or waving. The child uses his senses, such as vision, hearing, or touch to attend to or act on the object or event and the behavior is repeated many times. With representational play the child demonstrates his or her knowledge of how an object is used, the behaviors observed then are dependent on the object being used. For example, the child drinks from a cup or brushes his or her hair with a comb. Finally, the child's play becomes symbolic wherein, the child's actions are governed less by the presence of a given object and more by the child's own mental operations. We begin to see evidence that the child has a prior mental plan for his play. He is developing his own play schemes. For example, the child might substitute one object for another, such as a stick for a spoon. In the child's mental plan then stick equals spoon and this image then directs the child's actions on the stick. So that "stir with spoon" equals "stir with stick". Another example of this mental planning or exhibition of an intent to act is that the intended behavior is first verbalized. So the child verbally describes his intended play schemes. Another indication that a child has an internal plan

is when he verbalizes his intention and then actively searches for an object to be used in the execution of that plan. For example, the child may be holding a doll, say "baby thirsty", look for a cup, and then give the doll a drink. Lastly, a child could treat a doll or other object as an agent capable of performing actions independently. For example, rather than the child placing the bottle to the doll's mouth, the doll is made to hold the cup. It therefore becomes an active participant in the play. At the symbolic play level, the child's play becomes sequenced in a more logical order and the relationships between objects and agents becomes more sophisticated. The play scheme contains many combinations of actions on agents and objects. The child begins to generalize behaviors across objects and actions.

The Play Assessment Scale is a measure which allows an examiner to observe these hierarchical levels of children's play behaviors. It contains descriptions and exemplars of 45 play behaviors ranging from the two month age range to the 36 month age range. The behaviors are sequenced by 3 month intervals, for example, 2-4 months, 5-7 months, 8-10 months,

and so forth. The scale is object oriented and not meant to be used as a tool for judging social or peer interactions. It gives the examiner an indication of the child's knowledge of his world in the context of his or her actions on objects. The fact that the observer does not direct the child's behavior makes it a natural, unobtrusive means to obtain information about a child's spontaneous actions on toys and objects. Because play occurs across all environments, with many or no objects, alone or with others, with age level peers or adults, with familiar people or strangers and even with the absence of a common language, it seems to be an important medium for obtaining information on what and how to teach a child. An examiner can gather information on a child's toy preference and on the best level of prompting for a given child. The format is based on following the child's interest, unlike formalized tests which elicit only a given response and often provide no information of how the child learns.

Play is an important link to understanding a child's overall development as it is the primary means by which

children learn. The PAS may generate useful information for the planning of educational activities designed to promote skills in the areas of problem solving, language, and fine motor development. Results from the PAS may be used as a means by which to assign children to various play groups.

Play is an important tool by which a normal child grows and develops (Bruner, Jolly, and Sylva, 1976; Millar, 1968; Piaget, 1951) (cited in Li, 1981). Recently attention has been focused on the play of children with handicaps. Some limited information

regarding the characteristics of play of children with handicaps and the value of play for those children has been accumulated. This is the next topic of discussion.

Literature on play of children with handicaps

Research studies have been conducted which describe the play characteristics of children with various handicapping conditions including mental retardation, language impairments, autism, visually impairments, and hearing impairments.

Studies examining the play skills of mentally handicapped

children have typically compared the play skills of this population with normally developing peers. Wing, Gould, Yeates and Brierley (1977) studied the symbolic play of a population of severely mentally retarded children of 5 to 14 years of age. They found that varied and flexible symbolic play was only seen in children with a nonverbal mental age and receptive language ages of above 19 months. Weiner and Weiner (1974) examined the toy play characteristics of a group of 6-year-old mentally retarded children and two control groups of non-mentally retarded children. One of the control groups was matched for mental age and the other was matched for chronological age. The six year old mentally retarded subjects were found to demonstrate less sophisticated play patterns than the three year old children with whom they were matched for mental age. Hill and McCune-Nicolich's (1981) observation of play behaviors in children with Down syndrome, support the previous finding that mental age is more related to symbolic play levels than is chronological age. The children with Down syndrome in this study exhibited very few single, self-pretend play behaviors. A study by Whittaker (1980; cited in

Langley, 1985) found that profoundly mentally handicapped subjects exhibited play behaviors of self-feeding and self-brushing at a later developmental age than expected. In addition, these subjects had significant delays in combinatorial symbolic behaviors where a single scheme is applied to several agents (feed self then feed doll). These children did not make the transition from self-related to doll-related behaviors.

Tilton and Ottinger (1964) studied the play of autistic children. These autistic children spent a higher percentage of time in repetitive motoric manipulations and mouthing of toys, whereas normal children spent more time in combinatorial symbolic play.

Reynell (1978) found visually impaired children first begin to diverge from the developmental patterns of sighted children at about 10 months of age, "when perceptual characteristics guide the child to form conceptual relationships", in other words when sighted children are first beginning to associate objects with their functional use. Rogers and Puchalski (1984) found the visually impaired

children in their study were able to demonstrate symbolic play at a mean age of 25.9 months which is much earlier than the literature suggests. Interestingly enough, they report that use of the word "no" and two-word combinations signified readiness for symbolic play in visually impaired children.

Belsky and Most (1981) found that hearing-impaired children had a tendency to use toys as intended or representationally rather than as substitutes for other objects in symbolic play. These children also exhibited parallel play more frequently than interactive play. Also there is much evidence that hearing-impaired children's primary play deficits are related to communication, social, and symbolic relationships (Darbyshire, 1977; Higginbotham, Baker, and Neill, 1980; cited in Langley, 1985).

Relationship of play to language development

The relationship between play and language development has been studied extensively since Piaget (1962) first described children's progression from early sensorimotor play to symbolic play. Piaget stated that "play and language

development reflect the young child's ability to manipulate symbols". He proposed that play and language develop from the same representative or "semiotic" function. For both symbolic play and language to emerge the child must possess a "mental representation" of objects, people and events in his/her environment. The relationship between play and language then would be directly correlated, meaning that achievements in one domain should parallel achievements in the other domain. Other theorists have further hypothesized that the relationship is causal and the achievements in symbolic play should proceed linguistic development since representation should first be evidenced in the more concrete modality. Investigators have sought to lend support to each of these theories by pairing language-impaired children with normal children and then measuring play skills to note if any differences exist.

When children were matched on the basis of chronological age, the results supported the view that language-impaired children also exhibit impairments in their ability to play symbolically (Lovell, Hoyle, and Siddal, 1968; Brown, Redmond, Bass, Liebergott and Swope, 1975; cited in Terrell, Schwartz,

Prelock, and Messick, 1984). These investigations support the hypothesis that language and play develop in a parallel manner from some common base of knowledge dealing in mental representation.

When subjects were matched according to linguistic skills the language-impaired children demonstrated more sophisticated symbolic play than language skills in comparison to the younger control group subjects (Roth and Clark, 1987). These authors claim that based on this information, language and play may emerge from the same cognitive mechanism, but symbolic deficits are not necessarily uniform across both language and play. These results show that symbolic play skills can exceed the child's linguistic skills.

Terrell et al. (1984) found her subjects demonstrated impairments across both linguistic and symbolic play domains when they were matched with younger subjects who had similar linguistic skills. The language-impaired children in this study demonstrated play skills which were more advanced than their language skills and, their play skills were more advanced than the language-matched younger children's play

skills. However, the language-impaired subject's play skills were still below the level expected for their chronological age. These results then suggest that although some aspects of language production and play seem to be independent, an impairment of language appears to be concomitant with a similar though not equivalent ... deficit in symbolic play. Terrell et al. (1984) suggest that these findings support the hypothesis that rather than being directly or causally related, play and language interact in a "reciprocal" fashion. This would imply that when language deficits exist, the development of play may be limited or constrained.

Both the language-impaired and some of the younger normal subjects in this study demonstrated relationships in their play which were not yet "coded linguistically" meaning that "knowledge and concepts used in play are not translated directly into verbal expression". Therefore these children might demonstrate the knowledge of the coordination of the concepts agent + action + object (man drive tractor or doll cook dinner), but yet not verbally produce the same combination. Terrell et al. (1984) suggest that assessing play

skills may be directly useful in intervention procedures by directing therapy at language goals which express meanings the child spontaneously demonstrates in play schemes. Thus play scales may be used to assess language potential.

A previous study by Kahmi (1981) is consistent with the results obtained by Terrell et al. (1984). He found that the nonlinguistic or symbolic play behaviors of his language-impaired subjects were significantly higher than those of normal subjects matched on the basis of mean length of utterance. However, when he compared the language-impaired subjects' nonlinguistic symbolic abilities to a control group matched on the basis of mental age, the language-impaired subjects' performances were significantly lower. These findings support the notion that language deficits may be the result of not only a deficit in representational thought, but also due to some specifically linguistic skill.

McCune-Nicholich (1981) states further research is needed to understand the relationship that exists between play and language development. She suggests, "concurrent measures of play and language can be used to study the relationship

between these systems during their period of rapid development." It can be seen from the above mentioned literature review that in general, theorists support the notion that a relationship does exist between play and early language development, but that the exact nature of that relationship is as of yet undefined and is quite possibly some combination of all of the above mentioned theories and of theories yet to be proposed.

Each of these authors, however, uses some measure of play development to describe the play abilities of their subjects. These include; the Symbolic Play Test (Lowe and Costello, 1976), observations based on another author's descriptions (Piaget, 1962; Jeffree and McConkey, 1975; McCune-Nicholich, 1981), or a symbolic play task such as the "object stimulus gradient" developed by Casby and Ruder (1983) which was used to determine the extent to which objects could become dissimilar in children's symbolic object substitutions (one object represents another).

The purpose of this study is not to lend support to one or the other of the above mentioned hypothesizes of how

language and play development are related. Rather, this study is intended to examine that relationship by calculating concurrent validity measures between the Play Assessment Scale and the Battelle Developmental Inventory. Concurrent validity coefficients indicate the degree of relationship between "an instrument and a criterion measure which is assumed or known to be a reliable and valid measure itself" (Newborg et al. 1984). The BDI reports concurrent validity measures between the BDI and 7 other developmental assessment procedures. Newborg and colleagues (1984) go on to state, "Overall, the correlations between the BDI and the Vineland, DASI and Stanford-Binet offer strong support for the concurrent validity of the BDI."

On the other hand, no data is reported for concurrent validity between the PAS and any other assessment procedure. The current study will compare results obtained using the PAS and the Communication and Cognitive domains of the BDI in order to calculate concurrent validity coefficients and the study will also calculate interobserver reliability for each of these measures.

The rationale for conducting this study is to provide information regarding the validity of the PAS which may be used in programs designed to enhance the development of young children with handicaps. Results from the PAS might be used to substantiate results from other diagnostic assessment procedures, determine goals, identify toy preferences for intervention, and evaluate intervention programs. A limited amount of information concerning the accuracy and reliability of the PAS has been collected.

One study being conducted at the University of Utah, Early Intervention Research Institute, (C. Weber, personal communication), has preliminary data (pilot study of 9 subjects) investigating the concurrent validity of the PAS with all the domains on the BDI. Weber's results indicated moderately high correlations between the PAS and the BDI, in the domains of communication and cognition (.6734 and .7090 respectively). Correlations between the PAS and BDI personal-social, adaptive, fine motor, and motor total ranged from .4622 to .5541. The correlation between gross motor and the PAS was considerably lower (.0540).

Pagnotta (1988) also presented results from her graduate thesis. She examined the relationship between the PAS and an assessment procedure, the Alpern Boll for 25 subjects. Results indicated high correlations between the age levels obtained on cognitive and language sections on the Alpern Boll and the PAS play age.

Purpose of the study

The purpose of this study is to measure the concurrent validity between results obtained on the Play Assessment Scale (PAS) and the Battelle Developmental Inventory (BDI) (Cognitive and Communication Domains) using both normally developing children and children with handicaps. In addition, interobserver reliability was calculated. Specifically, this study addressed the following questions:

- 1) How does the PAS correlate with the BDI in terms of concurrent validity? Are the results obtained on the PAS similar to the results obtained on the Cognitive Domain of the BDI, the Communication Domain of the BDI, or both, if the assigned "play

age" (in months) is compared to the age equivalents (in months) obtained on the BDI?

- 2) If two examiners independently observe and score play behaviors on the PAS, are those results statistically significant in terms of interobserver reliability?
- 3) If two examiners independently score items obtained on the Cognitive and Communication Domains of the BDI, are those results statistically significant in terms of interrater reliability?

Chapter II

METHODS

Subjects

The subjects in this study were 18 children ranging in age from 4 months to 53 months. Subjects were excluded from the study if their scores exceeded the 36 month ceiling on the PAS. The mean age of the participants was 25.5 months. The subjects who participated in this study attended several local daycare centers in Missoula, Montana. These daycare centers were involved in a grant project designed to integrate children with handicaps into existing daycare homes. The sex of the subjects included ten boys and eight girls. The percentage of the subjects who had identified handicaps such as mental retardation, physical impairments or were determined "at-risk for a developmental delay" was 38. The remaining 61.1% of the subjects were considered to be normally developing children. The age, sex, and handicapping condition of each child are included in Table I.

Table 1. Individual subject characteristics.

Subject	Age	Sex	Handicapping Condition
1.	15m	F	*
2.	29m	F	*
3.	25m	F	*
4.	31m	F	*
5.	6m	F	*
6.	4m	F	*
7.	13m	M	*
8.	49m	F	MR
9.	11m	M	*
10.	44m	F	Spin-Bif., assoc. mild-mod.delays
11.	35m	M	At-risk DD
12.	51m	M	Down Syn., assoc. mod. delays
13.	37m	M	Seiz. Dis. At-risk DD
14.	6m	M	*
15.	15m	M	*
16.	30m	M	At-risk DD
17.	40m	M	At-Risk DD
18.	18m	M	**

Note. (*) represents not applicable, (M) represents months. Handicapping conditions = mental retardation, spina-bifida & associated mild-moderate delays, at risk for developmental delay, Down Syndrome with associated moderate delays and seizure disorder.

Procedures

All subjects were administered the Battelle Developmental Inventory and the Play Assessment Scale. Only the Cognitive and Communication Domains of the BDI were used in the correlational analyses with the PAS.

Each subject was seen and tested in the daycare setting. Two examiners administered the assessment protocols. The same examiner, however, administered both the BDI and the PAS to a specific child. The time between administration of these two measures did not exceed two weeks. The administration of the BDI and the PAS was counterbalanced for order of presentation to ensure that administration of one measure did not affect the outcome of the other. The two examiners were professionals in early intervention with whom the children were reasonably acquainted. The examiners were experienced in the assessment of infants and preschool children with handicaps and were qualified to administer both the BDI and PAS based on the specifications provided in the assessment manuals. In addition, both examiners had attended workshops intended to

train potential users on administration, scoring, and use of both these instruments.

Prior to obtaining the reliability data for the PAS, the examiners discussed and agreed upon an item by item interpretation and scoring procedure. As neither had previously administered the PAS, they practiced administration and scoring procedures based on observations of two children. Following the PAS training a point by point reliability percentage of .733 was obtained on 3 children not included in the reliability sample. The examiners then discussed specific items on which they disagreed.

The administration of each assessment procedure was videotaped using a Panasonic-Color Video Camera, WV-300. Interrater reliability was computed on 22% (N=4) of randomly selected administrations of the BDI assessments completed. Interrater reliability was determined with one examiner scoring the Cognitive and Communication Domains of the BDI during administration of the items and a second observer scoring the items from a videotape. Interrater reliability was determined on 55% (N=10) of randomly selected administrations

of the PAS assessments completed. The reason for calculating the lower percentage (22) of the BDI assessments in order to determine interrater reliability was because the manual reported excellent interrater reliability for each of the domains. In addition, a recent study by McLean, McCormick, Bruder and Burdg (1987) reported .908 and .940 interrater reliability on the Communication and the Cognitive Domains of the BDI respectively. On the other hand, the Play Assessment Scale did not report any reliability measures nor current research to support its reliability. Therefore the higher 50% criteria was used to ensure an appropriate degree of interrater reliability.

Chapter III

RESULTS

The first research question addressed by this study was-- what is the relationship between the Battelle Developmental Inventory (BDI) Cognitive and Communication Domains and the Play Assessment Scale (PAS) in terms of concurrent validity? The second research question was to determine what the percentage of agreement would be for interrater reliability for the BDI and PAS. For the purposes of this investigation the following analyses were conducted on the BDI and PAS:

- correlational analysis of the PAS "play age" and the BDI, Cognitive and Communication Domain "age equivalent".

- interrater reliability on the BDI, Cognitive and Communication Domains.

- interrater reliability on the PAS

The Pearson Product Moment Correlation Coefficient was used to determine the relationship between the BDI "age

equivalents" and the "play ages" obtained on the PAS for each of the subjects included in this study. The results of this relationship are reported and presented in tabular form (Table 2). The results to assess concurrent validity will be presented first, followed by results regarding interrater reliability for first the BDI, Cognitive and Communication Domains and then the PAS.

Concurrent Validity

The Pearson Product Moment Correlation Coefficient (Friedman, 1986) was applied to the pairings of "age equivalents" of the BDI Cognitive and Communication Domains and the "age equivalents" of the PAS. The results of these comparisons are presented in Table 2. The correlations ranged from .92788 to .93694 and were significant at the .01 level of confidence. The degree of correlation for each domain was high and supports the concurrent validity of the PAS in relation to both the Cognitive and Communication Domains of the BDI. These results indicate that an examiner can obtain similar information concerning a child's developmental levels

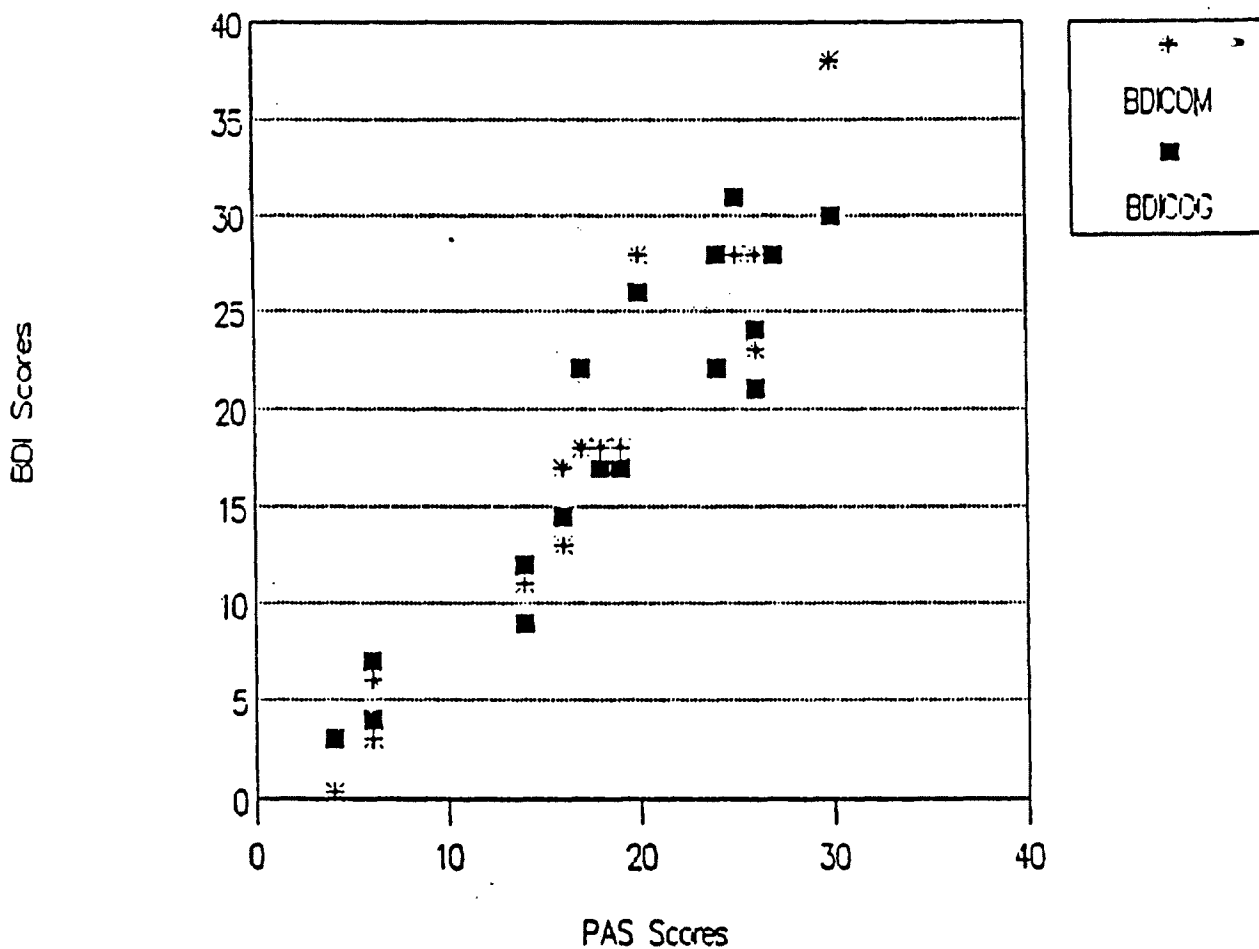
using the PAS or the BDI Cognitive and Communication Domains. Correlations between the Cognitive and Communication Domains, although not related to this study, are presented in Table 2 as well. Figure 1 shows the scattergram of the play age scores obtained on the PAS as a function of the BDI Cognitive and Communication age equivalent scores.

TABLE 2. Correlation coefficients between the "play ages" of the Play Assessment Scale and the Age Equivalencies of the Battelle Developmental Inventory Cognitive and Communication Domains.

	BDI COGN	BDI COMM	PAS
BDI COGN	1	.93239	.92788
BDI COMM	.93239	1	.93694
PAS	.92788	.93694	1

Note. BDI COGN represents Battelle Developmental Inventory Cognitive Domain, BDI COMM represents Battelle Developmental Inventory Communication Domain, and PAS represents Play Assessment Scale.

Figure 1. Scattergram of PAS "play age" scores as a function of BDI "age equivalent" scores.



Note: PAS refers to Play Assessment Scale, BDICOM refers to Battelle Developmental Inventory, Communication Domain, and BDICOG refers to Battelle Developmental Inventory, Cognitive Domain.

Interrater Reliability

Interrater reliability was obtained on 22% of the BDI assessments (n = 4) and 55% of the PAS assessments (n = 10). A point by point interrater reliability calculation was determined using the following formula (McClellan et al., 1987):

$$\frac{\# \text{ of agreements}}{\# \text{ of agreements} + \# \text{ of disagreements}}$$

As noted from the data presented in Table 3, interrater reliability was high for the BDI Cognitive and Communication Domains. Agreement was well above 90% across both domains. The combined agreement was .961 on all administered items, demonstrating a high degree of interrater reliability. Based on this data, two examiners are likely to similarly score items on the BDI for these two domains.

Table 4 displays the interrater reliability data for the PAS. Agreement was moderately high at 77%. This result indicated that two examiners were less likely to score items in the same manner on the PAS as compared to reliability data for the BDI.

Table 3. Interobserver reliability measures for the Battelle Developmental Inventory Cognitive and Communication Domain

	Cognition	Communication	Combined
# of Agreements	52	46	98
# of Agreements + # of Disagreements	55	47	102
Percentage of Agreements	.945	.979	.961

Table 4. Interobserver Reliability Measures for the Play Assessment Scale

	<u>PAS</u>
# of agreements	77
# of agreements + # of disagreements	100
Percentage of Agreements	.77

Note. (PAS) represents Play Assessment Scale

Chapter IV

DISCUSSION

Concurrent validity of the Play Assessment Scale (PAS) and the Battelle Developmental Inventory (BDI) will first be discussed with regard to theoretical issues and then with regard to clinical application." In the second section, theoretical aspects and the clinical applications of the reliability results will be discussed. Finally, the limitations of the present study and implications for future research will be presented.

Concurrent Validity

One purpose of the present study was to determine the relationship between the PAS and the Cognitive and Communication Domains of the BDI in terms of concurrent validity. The results indicated that a high correlation existed between the "play age" obtained on the PAS and the cognitive and communication domain "age equivalents" obtained

on the BDI. Concurrent validity studies, (McClellan, McCormick, Bruder and Burdick, 1987; Guidubaldi, Newman, Cleminshaw, Perry, Telzrow, Serazin, Maranda, Vettel and Harr, 1981; and Newborg, et al., 1984), have compared the BDI with frequently used assessment procedures such as the Bayley Scales of Infant Development (Bayley), (Bayley, 1969), Vineland Adaptive Behavior Scales (Vineland), (Sparrow, Balla, and Cicchetti, 1984), Vineland Social Maturity Scale, (Vineland), (Doll, 1965), Developmental Activities Screening Inventory (DASI), (Dubose and Langley, 1977), Stanford-Binet Intelligence Scale (Stanford-Binet) (Terman and Merrill, 1960), Wechsler Intelligence Scale for Children - Revised (WISC-R), (Wechsler, 1974) and Peabody Picture Vocabulary Test (PPVT), (Dunn and Dunn, 1981). The DASI is used by professionals to screen cognitive and perceptual skills of children. Professionals use the following tests to diagnose and qualify children for intervention services: Stanford-Binet, WISC-R, and Bayley (measures of intelligence); Vineland, (measures adaptive and personal-social skills) and PPVT-R (measures receptive vocabulary). Results of the above mentioned studies supported

the concurrent validity of the BDI. The correlations between the BDI and the DASI (.78 to .92) (Newborg et al., 1984), Vineland (.79 to .94) (Newborg et al., 1984) and the Bayley (.75 to .92), (McClellan et al., 1987) were all high in these studies. The BDI also correlated positively with the Stanford-Binet (.40 to .61) (Newborg et al.,) and PPVT-R (.36 to .83) (Newborg et al., 1984). The correlations between the BDI and WISC-R were less positive (.02 to .79), although the authors (Newborg et al., 1984) stated these results were obtained on a small number of subjects. These tests, with the exception of the Vineland, require structured, standardized administration and scoring procedures as does a significant portion of the BDI. These findings supported the psychometric integrity of the BDI. The BDI, then, is presumed to be a good criterion measure to determine the relationship between the PAS and a structured, standardized assessment procedure. The results of the present study indicated a high correlation between play as measured by the PAS and cognitive and language abilities as measured by the BDI. The implications of these results for clinical use are that professionals can obtain

similar information using the PAS in comparison with the BDI. Furthermore, these results lead to the inference that an unstructured, observational assessment procedure such as the PAS generates similar information relative to a child's level of functioning as the above mentioned structured test procedures yield.

Psychometric integrity of preschool measures is a primary concern within the field of early education (Brookes-Gunn and Lewis, 1981; Johnson, 1982; Sheenan, 1982; Bracken, 1987; Sheenan, 1989; Bailey and Bricker, 1986). Sheenan (1989) states "that less than 10% of the existing preschool assessment tools are thought to have adequate reliability and validity". This is a vast number when the number of possibilities for assessment procedures "exceeds 300". A related issue is that there are limited assessment procedures which allow examiners to rely on their observational skills in a systematic manner (Niesworth and Bagnato, 1989). Assessment procedures which allow examiners to use observational skills, but also meet psychometric requirements are an important addition to the field. The results of this

study support the psychometric integrity of the PAS in terms of concurrent validity. This study indicated that for the population of children under 3 years of age, an observational, unstructured procedure provided similar information, with respect to age levels, as a structured test. This has vast implications for the clinical use of the PAS, which will be discussed later.

As was discussed in Chapter I, many problems exist in assessing young children with special needs (Brookes-Gunn and Lewis, 1981; Johnson, 1982; Fewell, 1983; and Bailey and Worley, 1984). For many young children with handicaps, performance on a standardized test may not provide a representative sample of a child's actual abilities due to interferences such as, noncompliance, distractibility, sensorimotor impairment, shyness or motivation. Many standardized tests are not readily adapted for hearing-impaired, visually-impaired or motorically-impaired children. They also may not be suited for emotionally disturbed or autistic children.

The PAS may help resolve some of the problems associated

with assessing "special needs" preschoolers. Evaluators may be less likely to encounter withdrawals, task refusals, behavior outbursts and general noncompliance when using the PAS. Therefore, the PAS may provide a more representative sample of child behavior if such behaviors interfered with obtaining assessment results when using a structured procedure. Also, if the PAS were used in conjunction with other assessment procedures, it could be used to help interpret the results of structured assessment procedures, especially given the high correlations between play and cognition and communication obtained in this study. Given the case of a child who does not, for some reason or another, perform well on structured intellectual or language tests, PAS results might actually more accurately reflect his/her abilities than the structured test. For example, if a child's functioning age levels were lower on standardized measures of intelligence or communication than on the PAS, an assumption could be made that the child had not performed up to his abilities on the structured tests or that the structured test did not accurately represent the child's abilities.

The PAS has applicability with respect to the implementation of Public Law 99-457, the Education of the Handicapped Act Amendments. The field is in need of valid and reliable measures to be used in assessing the population, birth to 2 years, which are affected by this law (Sheenan, 1989). The current findings regarding the concurrent validity of the PAS, indicate that it may aid professionals in the implementation of P.L. 99-457.

Under P.L. 99-457, children, age birth to 2 years, do not need a categorical label to receive services, as do older school age children. Children who are considered "at-risk" for a developmental delay based on such factors as medical, biological or environmental factors can receive special services under this law. The PAS then could be used as part of a qualifying "package" of measures to determine eligibility for services. Under P.L. 99-457, an Individualized Family Service Plan (IFSP) will be developed for each child and family to receive services. The IFSP will document outcomes and objectives for the child and family. Assessment and ongoing monitoring of progress becomes an important

consideration. The PAS could conceivably be used to monitor child progress when intervention is aimed at increasing environmental stimulation, parent-child interaction or stimulation based on peer interaction when a child has been placed in a daycare or preschool setting as an objective. Use of the PAS then could document program effectiveness based on behavioral changes or skill development across any of the above mentioned dyads. Considering the results obtained regarding concurrent validity, the PAS might document child progress in cognitive or communication intervention programs. Based on these results, an assumption can be made that changes in communication and problem-solving skills will be reflected through changes in play skills.

As suggested by Johnson (1982) and Fewell (1983), alternative assessment procedures are needed which will offer information in addition to information standardized tests yield, particularly when placement and intervention are being considered. The results of the present study indicate the PAS is a viable alternative assessment procedure which may offer additional information. The results of the present study also

suggest that observational assessments can be valid instruments. The high concurrent validity obtained in the current study should alleviate the concern of Bailey and Worley (1984). Specifically they were concerned that observational procedures lacked guidelines allowing consistent interpretation by various examiners. The PAS guidelines appear to be adequate given the degree of concurrent validity demonstrated between the PAS and the BDI.

Another application of the PAS would be to include it in a multidimensional assessment approach (Niesworth and Bagnato, 1988). In this approach information is gained from a variety of sources and measures including observations or clinical judgments across a variety of dimensions. Because the current study resulted in high concurrent validity for the PAS, it could be used in a multidimensional approach. Considering these results, examiners are assured that they are using a measure which correlates highly with a structured assessment procedure, at least in cognition and communication domains.

The findings of the current study resulted in concurrent

validity correlation coefficients which were higher than those reported by Weber (undated). Her correlations between the PAS and the cognitive domain of the BDI were .6734 and the communication domain of the BDI were .7090, in comparison to the .9279 and .9324 obtained respectively in the current study. Weber's data was collected using only 9 subjects as compared to the 18 used in the present study which may account for some differences found between the two studies. Furthermore, Weber's observations of play behavior were conducted as her subjects played with their mothers, which may have influenced the results. The children in the present study played in the presence of professionals. The children in Weber's study had handicapping conditions, whereas only 38% of the subjects in the current study had handicapping conditions. This may indicate that the PAS is not as sensitive when used to assess the play skills of children with handicapping conditions. Further research assessing the relationship between standardized tests and the PAS with children with handicapping conditions may would offer more information regarding this issue.

Relationship between play, language, and cognition

In Chapter I, the relationship between cognition, play and language development was discussed. The literature seems to support the notion that cognition and play are closely related. Studies investigating the play of children with mental retardation have indicated that play is more closely related to mental age than chronological age (Weiner and Weiner, 1974; Hill and McCune-Nicolich, 1981; Whittaker, 1980). Researchers have also been interested in the relationship between cognition and language. They have been particularly interested in this relationship as it relates to representational thought. Thus, the interest in the correlation between language development and play development. The results of the current study indicated that play as measured by the PAS and language as measured by the BDI Communication Domain are closely related. The high correlation of .93239 supports the hypothesis that play and language development are related to each other. Also the high correlation (.92788) between play skills and cognitive skills

suggests that a child's play skills may be a manifestation of his /her cognitive skills. These results support the hypothesis that a child's play skills are a reflection of an underlying component related to representational thought. As a child's understanding of the world increases and his or her knowledge about the relationships which exist between objects and people becomes more sophisticated, concurrently his play and language development will reflect these changes. Play with objects will become less sensory oriented and more functional. For example, rather than place a comb in his /her mouth a child will comb his/her hair. The child will substitute similar objects for objects used in play and play schemes will become longer and more complex. As these changes occur in play, concurrent changes will occur in language. A child will be learning names for familiar people and objects as he is beginning to use play to represent his understanding of how objects are used. He will begin to combine words into short phrases as he demonstrates combinations or sequences of play schemes (McCune-Nicolich, 1981). The concurrent validity results of this study support the idea that play and language

development are parallel and are closely related to cognitive abilities. There are a number of hypotheses regarding what the exact nature of the relationship between play and language development maybe. The purpose of this study, however, was not to determine what that relationship might be (for example, "causal" or "reciprocal"). Researchers examining the relationship between play and language might, however, use the PAS as a way to measure play skills in future studies.

Reliability

The results of the present study indicate the PAS is a fairly reliable assessment procedure. Interobserver reliability was calculated to be 77% between two examiners using a point by point analysis. Caution must be used in interpreting these results. In preparing for this study, the examiners noted some difficulties with the PAS which may have effected the degree of interobserver reliability. The following concerns prompted the need for further training to occur:

Some of the play behaviors to be observed on the PAS can be displayed only momentarily.

Behaviors need only be observed once in order for the child to receive credit.

Some of the items are discretely different from one another, but not sequenced one after another making scoring somewhat difficult for new examiners.

The manual provides descriptions of behaviors to be observed and several exemplars of these behaviors. Despite this, the examiners in this study found there was room for individual interpretation on many items. Disagreement might arise concerning whether or not an observed behavior was a true example of a PAS item.

The degree of reliability obtained in this study would most likely not have reached the level it did without the implementation of the following preliminary training and preparation.

Reliability Training

Although the two examiners participating in this study had no previous experience in administration of the PAS, they did attend a workshop which instructed potential examiners in the administration and scoring of the PAS. In addition, the examiners discussed and agreed on the interpretation and scoring procedure for each item prior to administration of the PAS. They then practiced on several children not participating

in the study and established the pre-reliability score of .73 on 3 subjects not included in the reliability sample. The reliability information presented in this study was most likely influenced by this degree of preparation. Interestingly enough, the examiners did establish a reliability level of .73 following the item by item discussion and the degree of reliability only increased by 4% after the play of 3 children was observed using the PAS. The item by item discussion and interpretation may have had more affect on the percentage of interobserver reliability than did the practice reliability sessions. However, the exact degree to which the discussion and interpretation effected the results is not known since the percentage of agreement prior to the reliability training was not established.

The degree of interobserver reliability obtained in this study impacts the clinical use of this procedure. The examiners in this study had to prepare and train to establish the moderate degree of reliability reported in the results section (.77). These results indicate that it is important to establish some measure of reliability prior to using the

PAS. The moderate interobserver reliability results obtained on the PAS are likely related to the fact that the PAS is less structured and therefore less likely to reach the degree of reliability that can be found with structured tests. For example, the degree of reliability for the cognitive and communication sections of the BDI combined was .961. The PAS play behaviors and exemplars are somewhat ambiguous and therefore more subjective when compared to the BDI which provides standard procedures for scoring responses. Compared to other similarly structured assessment procedures such as, the Uzgiris-Hunt Scales of Infant Psychological Development, (U-Z), (Dunst, 1980), the interobserver reliability percentage on the PAS is still lower. Dunst (1980) reports that "the percentage of agreement between independent observers has generally been in the 0.85 to 0.99 range". Considering the lower reliability findings of the present study, this author suggests a similar training procedure to the one described in this study be used by PAS observers to ensure that an adequate degree of reliability is established. An important element of the training is for examiners to

discuss and to agree upon an item by item interpretation and scoring procedure, this element resulted in the .73 agreement obtained prior to actually gathering the reliability data.

Considering the reliability results, caution is advised in terms of Fewell's statement in the manual that examiners can be parents or other familiar adults. The moderate degree of reliability obtained by experienced observers would suggest that untrained observers should not score the PAS. Early education specialists recognize the need for and the importance of parent participation in the assessment process, however, caution is advised in allowing parents to administer and score the PAS. Instead the parent could be perhaps be involved in the administration of the PAS on a less direct level. They could sit near the child and be instructed to respond only if the child initiated interaction. They could also be coached in terms of which behaviors to prompt.

The PAS would be a useful assessment procedure in a transdisciplinary team approach in which team members rely on each other to elicit information relative to a specific discipline. Team members could gain information relative to

the areas of cognition, fine motor, adaptive, and communication and to a lesser degree gross motor. by using the PAS. Any team member could administer the PAS, if trained. If the PAS were to be used in conjunction with other formalized procedures, it would be an excellent vehicle for "warming" a child to a test situation and unfamiliar clinicians. Useful information would be obtained during the "warm-up" time making the PAS a cost effective procedure. Added to this is the relatively short administration time of between 20 minutes and less than one hour. The same cautions apply with respect to the interobserver reliability findings found in this study. A diagnostic team should demonstrate reliability among the team members prior to using the PAS as a clinical tool.

The question arises, however, concerning reliability between agencies and individuals wherein establishing reliability is not feasible. On the one hand, the results of this study indicated high concurrent validity between the PAS and a structured assessment procedure, but some degree of reliability has been sacrificed for the less structured design

of the PAS. Perhaps the degree of interobserver reliability could be increased making the PAS a more desirable assessment procedure. For example, clearer definitions and more exemplars of each play behavior to be observed could be provided. Specification of the toys and materials to be used when observing the play behaviors and videotapes describing and depicting behaviors representing specific PAS items would perhaps increase interobserver reliability. Furthermore, these suggestions would not affect the natural, unobtrusive design of the PAS, but they might increase the likelihood that independent examiners would score the PAS in a similar manner.

FUTURE RESEARCH

One of the limitations of the present study was the limited sample size, particularly with respect to children with handicapping conditions (N=7). This study had no visually- or hearing-impaired subjects and only one of the subjects had a severe motor impairment. The present study did not control for age as a factor in determining concurrent validity and reliability. This study examined only the

relationship between the "play age" and cognitive and communication "age equivalents".

To support the current findings, similar studies should be conducted to determine if these results can be replicated. Future studies might re-examine the relationship between the PAS and the BDI in terms of concurrent validity and also obtain scores for interobserver reliability. Similar studies of concurrent validity and reliability might be conducted using other norm-referenced tests such as the Bailey Scales of Infant Development or the Stanford-Binet. Research is needed to determine what the relationship between the PAS and standardized procedures is when these are used with multiply impaired, visually impaired, hearing impaired, motor impaired, as well as communication and cognitively delayed subjects, particularly in light of the discrepancy between the current results and those obtained in the pilot study by Weber (undated). As of yet, the PAS does not contain adaptations for these populations, although these are reportedly being developed. It would be interesting to investigate whether or not these adaptations affect the concurrent validity or the

percentage of interobserver reliability as reported by the current study and the previously mentioned studies which report these findings.

Future research should control for ages of the subjects to determine if the PAS varies in degree of reliability across differing age groups. In the present study reliability decreased as play ages increased. Age was not, however, one of the variables which was examined in the present study, so this would need to be documented with further research. For example, if subjects were divided into age groups, for instance from birth to 12 months, 13 months to 24 months and 25 months to 36 months would interobserver reliability results differ significantly? Correlations were not calculated for concurrent validity as a function of age for the subjects in this study. Would concurrent validity on the PAS vary significantly for subjects grouped according to the above suggested age ranges?

Future research might also be conducted to determine the test-retest reliability of the PAS. Would subsequent observations using the PAS yield the same results as initial

observations? This question is particularly important considering that an examiner can prompt behaviors not demonstrated by the child. Presumably, a child could learn a new play skill, but only if the examiner had modeled or verbally prompted the response.

Studies which assessed the applicability of the PAS in intervention programs would be useful. Specifically, studies examining early interventionist's perceptions of the PAS following administration and implementation of results would present information on its usefulness. For example, how would observers rate the PAS in terms of ease of interpretation and scoring of play behaviors, administration time, use with children with severe handicaps, or comparisons with other measures? Does the PAS provide information relative to a child's development which can be used on an IFSP or treatment plan?

Research is needed which would support the inference that the PAS correlates positively with other measures, based on the positive correlations which were found between the BDI and the Bayley, Vineland, Stanford-Binet and DASI. This research

is needed to more clearly define what uses the PAS might have in the assessment of children. Would it possibly predict developmental levels in other areas, such as adaptive skills, fine motor skills, or social skills? Research on the relationship between the PAS and other measures of play in terms of concurrent validity would further support the clinical use of the PAS. For example, how would the PAS correlate with the Symbolic Play Test (Lowe and Costello, 1976)?

Conclusion

In conclusion, the PAS is an important addition to assessment options available to professionals involved in the field of early intervention. The results of this study indicated that a high correlation exists between the Play Assessment Scale (PAS), (Fewell, 1986) and the Battelle Developmental Inventory (BDI), (Newborg et. al.) Cognitive and Communication Domains. The BDI is presumed to be concurrently valid in itself, based on high correlations between it and other well established assessment procedures. The importance

of these findings not only relate to the concurrent validity of a new assessment procedure, but to the concurrent validity of an unstructured assessment procedure. The PAS is a nonstandardized, nonintrusive procedure which allows examiners to use their observational skills in a systematic manner. Professionals have long relied on traditional structured tests to assess children's abilities. The correlations between the PAS and BDI suggested that the PAS can be used in combination with or in place of these traditional measures. The results of a moderate percentage of interobserver reliability indicated that means should be taken to increase the reliability of the PAS. Suggestions were offered which might result in higher interobserver reliability. Suggestions for establishing reliability prior to use of the PAS were also provided.

BIBLIOGRAPHY

- Bagnato, S. J., Neisworth, J. T., & Capone, A. (1986). Curriculum-based assessments for the young exceptional child; Rationale and Review. Topics in Early Childhood Special Education, 6 (2), 97-110.
- Bailey, E. J., & Bricker, D. A (1981). Psychometric study of a criterion referenced assessment instrument designed for infants and young children. Journal for the Division for Early Childhood, 4, 11-27, 1981
- Bailey, D. B., & Worley, M. (1984). Teaching Infants and Preschoolers with Handicaps. Columbus, Ohio: Charles E. Merrill.
- Bayley, N. (1969). Bayley Scales of Infant Development. New York; Psychological Cooperation.
- Belsky, J., & Most, R. K. (1981). From Exploration to play; A cross-sectional study of infant free play behavior. Developmental Psychology, 17, 630-639.
- Bracken, B. A., (1987). Limitations of Preschool Instruments and Standards for Minimal Levels of Technical Adequacy, Journal of Psychoeducational Assessment, 4, 313-326.
- Bricker, D.D. (1982). Intervention with At-Risk and Handicapped Infants. Baltimore: University Park Press.
- Brookes-Gunn, J. & Lewis, M. (1981). Assessing young handicapped children: Issues and solutions. Journal of the Division for Early Childhood, 3, 84-95.
- Casby, M. W., & Ruder, K. F., (1983). Symbolic play and early language development in normal and mentally retarded children. Journal of Speech and Hearing Research, 26, (3), 404-412.
- Doll, E. A., (1965). Vineland Social Maturity Scale, Circle Pines, Minnesota; American Guidance Service.
- DuBose, R. F., & Langley, M.B., (1977). Developmental Activities Screening Inventory. Hingham, Massachusetts: Teaching Resources Corporation.
- Dunn L. M., (1981). Peabody Picture Vocabulary Test-Revised. Circle Pines, Minnesota: American Guidance Corporation.

- Dunst, C. J., (1980). A Clinical and Educational Manual for use with the Uzqiris and Hunt Scales of Infant Psychological Development, Baltimore, Maryland: University Park Press.
- Fewell, R. R., (1986). Play Assessment Scale 5th Revision, available from the author, EEU WJ-10, University of Washington, Seattle, WA 98195.
- Garwood, S. G. & Fewell, R. R. (1983). Educating Handicapped Infants. Rockville, Maryland: Aspen Publishers, Inc.
- Hanson, M. J. (1984) Atypical Infant Development. Baltimore: University Park Press.
- Harrington,, R. (1985). Battelle Developmental Inventory. In D.J.Keyser & R. C. Sweetland (Eds.), Test Critiques, (Vol.II). Kansas City, MO: Test Corporation of America.
- Hill, P. M., & McCune Nicolich, L. (1981) Pretend play and patterns of cognition in Down's Syndrome children. Child Development, 52, 611-617, 1981.
- Jeffree, D. M., McConkey, R. (1976). An observation scheme for recording children's imaginative doll play. Journal of Child Psychology and Psychiatry, 17, 189-97.
- Kahmi, A. (1981). Nonlinguistic symbol and conceptual abilities in language-impaired and normally-developing children. Journal of Speech and Hearing Research, 24, 446-453.
- Langley, M. B., (1985). Selecting, adapting and applying toys as learning tools for handicapped children, Topics in Early Childhood Special Education, (3), 101-118.
- Lowe M. & Costello, A. J. (1976). Manual for the Symbolic Play Test. NFER, Windsor.
- Li, A. K. F, (1981). Play and the mentally retarded child. Mental Retardation, 19, (3).
- McCune Nicolich, L. (1981). Toward symbolic functioning: Structure of early pretend games and potential parallels with language. Child Development, 52, 785-797.

- McClellan, M., McCormick, K., Bruder, M., Burdick, N., (1987). NA investigation of the validity and reliability of the Battelle Developmental Inventory with a population of children younger than 30 months with identified handicapping conditions. Journal of the Division for Early Childhood 11, (3).
- Mott, S. E., Fewell, R.R., Lewis, M., Meisels, S.J. Shonkoff, J.P., & Simeonsson, R. J. (1986). Methods for assessing child and family outcomes in early education programs: some views from the field. Topics in Early Childhood Special Education, 6 (2), 1-15.
- Neisworth, J. T., and Bagnato, S. J., (1989). Assessment in Early Childhood Special Education: A typology of dependent measures. In Odom, S. L., & Karnes, M.B., Early Intervention for infants and children with handicaps, 23-49, Paul H. Brookes publishing Co.
- Newborg, J., Stock, J. R., Wnek, L., Guidubildi, J., & Svinicki, J. (1984). Battelle Developmental Inventory. Allen, TX: Teaching Resource.
- Pagnotta, P., (1988). Play Assessment of the young handicapped child: A comparison of play age and developmental age. Paper presented at Montana's 1989 State Conference on Establishing Optimal Services for children with handicaps.
- Piaget, J. (1962) Play, Dreams and Imitation in childhood. New York: Norton.
- Reynell, J., (1978). Developmental patterns of visually handicapped children. Child Care, Health and Development, 4, 291-303.
- Rogers, S. J., & Puchalski, C. B. (1984). Social Characteristics of visually impaired infant's play. Topics in Early Special Education, 3, 52-56.
- Roth, F. R., & Clark, D. (1987). Symbolic play and social participation abilities of language-impaired and normally developing children. Journal of Speech and Hearing Disorders 52, (1).
- Sheenan, R. (1989). Implications of P.L. 99-457 for assessment. Topics in Early Childhood Special Education, 8, 4, 103-115.
- Sparrow, S., Balla, D., & Cicchetti, D. (1984). Vineland Adaptive Behavior Scales. Circle Pines, MN: American Guidance Service.

- Terrell, B. Y., Schwartz, R. G., Prelock, P.A., & Messick, C.K. (1984) Symbolic play in normal and language-impaired children. Journal of Speech and Hearing Research, 27, 424-429.
- Terman, L. M., & Merrill, M. A., (1960). Stanford-Binet Intelligence Scale, Third Revision (form L-M). Boston: Houghton Mifflin Company.
- Tilton, J. R., & Ottinger, D. R. (1964) Comparison of the toy play behavior of autistic, retarded, and normal children. Psychological Reports, 15, 967-975.
- Wechsler, D., (1974). Wechsler Intelligence Scale for Children-Revised. New York: The Psychological Corporation.
- Weiner, E. A., & Weiner, E. J., (1974). Differentiation of retarded and normal children through toy-play analysis. Multivariate Behavior Research, 9, 245-252.
- Wing, L., Gould, J., Yeates, S. R. & Brierley, L. M. (1977). Symbolic play in severely mentally retarded and in autistic children. Journal of Child Psychology and Psychiatry, 18, 167-178.
- Zelazo, P. R., & Kearsley, R. B. (1980). The emergence of functional play in infants: Evidence for a major cognitive transition. Journal of Applied Developmental Psychology, 1, 95-117.