Validation of a Preschool Screening Measure: The Dial-R

Wendy Simmons-Watts
Western Kentucky University

Follow this and additional works at: http://digitalcommons.wku.edu/theses
Part of the Education Commons, and the Psychology Commons

Recommended Citation
http://digitalcommons.wku.edu/theses/951

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
VALIDATION OF A PRESCHOOL SCREENING MEASURE:
THE DIAL-R

A Thesis
Presented to
the Faculty of the Department of Psychology
Western Kentucky University

In Partial Fulfillment
of the Requirements for the Degree
Education Specialist in School Psychology

by
Wendy Carole Simmons-Watts
July 1994
VALIDATION OF A PRESCHOOL SCREENING MEASURE:
THE DIAL-R

Date Recommended 7/27/94

[Signatures]

Director of Thesis

[Signatures]

Lori E. Payne

[Signatures]

Director of Graduate Studies 8/23/94
ACKNOWLEDGEMENTS

Although this is an acknowledgement of the people responsible for helping this thesis become a reality, it is also for those who have helped me through the years it took to reach this point: my family, friends, and the faculty at Western Kentucky University have provided me with the support necessary to reach my goal of becoming a school psychologist; my parents who provided support and love, even when they did not understand; my husband, who was willing to stand aside from day one in order to let me achieve and offered encouragement when I faltered; my committee members, Dr. Elizabeth Jones, Dr. Shula Ramsay, Dr. Lois Layne, and Dr. John Bruni, who were flexible and provided their expertise and time -- above and beyond -- to help me succeed. Thank you to all.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ......................................................................................... iii  
LIST OF TABLES ................................................................................................... v  
ABSTRACT ............................................................................................................... vi 

TEXT  

I. INTRODUCTION .......................................................................................... 1  
   Early Intervention .................................................................................... 1  
   Assessment ............................................................................................... 4  
      Preschool Measures ............................................................................. 4  
      Technical Adequacy ........................................................................... 6  
      Instruments .......................................................................................... 9  
      Demographic Variables ..................................................................... 11  
   Rationale ................................................................................................... 16  

II. PURPOSE .................................................................................................... 19  

III. METHOD ................................................................................................... 22  
   Subjects .................................................................................................... 22  
   Instruments .............................................................................................. 22  
      DIAL-R ................................................................................................. 22  
      WPPSI-R .............................................................................................. 24  
   Procedure ................................................................................................. 27  
   Analysis .................................................................................................... 27  

IV. Results ....................................................................................................... 28  

V. Discussion .................................................................................................. 30  

APPENDICES ................................................................................................. 49  
   Appendix A Parent Permission Form ................................................... 49  
   Appendix B Evaluation Result Form ..................................................... 53  

REFERENCES ................................................................................................. 55  

iv
## LIST OF TABLES

1. Subject Demographic Characteristics ..................................... 41
2. Decision-Making Validity of the DIAL-R when Compared with the WPPSI-R on 54 Preschool Children ............................. 42
3. Correlation between the DIAL-R Area and Total Scores and WPPSI-R IQ Scores for a Sample of 54 Preschool Children ................................................................. 43
4. Mean, Standard Error of Measurement, and Standard Deviation for the DIAL-R and the WPPSI-R ............................................. 44
5. Mean Difference between the DIAL-R Total Score and the WPPSI-R Full Scale IQ ................................................................. 45
6. Frequency of the Difference between the DIAL-R Total and WPPSI-R FSIQ ................................................................. 46
7. Mean, Standard Error of Measurement, and Standard Deviation for the DIAL-R and the WPPSI-R Total Scores for Each Demographic Variable ............................................. 47
8. T-test Scores for Mean Differences on the DIAL-R and WPPSI-R Total Scores for each Demographic Variable ................................................................. 48
VALIDATION OF A PRESCHOOL SCREENING INSTRUMENT:
THE DIAL-R

Wendy Carole Simmons-Watts  July 1994  61 Pages
Directed by: Elizabeth Jones, Shula Ramsay, Lois Layne, and John Bruni
Department of Psychology  Western Kentucky University

The classification and correlational validity of the DIAL-R was investigated for an at-risk population of preschoolers. Fifty-four preschool children (mean age 60.87 months) were administered the Developmental Indicators for the Assessment of Learning-Revised (DIAL-R) and the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R). The classification agreement between the DIAL-R and WPPSI-R was high (.89) and statistically significant (Chi-Square(1) = 19.01); however, the DIAL-R failed to identify six children in need of services resulting in a sensitivity index of .40. The DIAL-R did not overidentify any children resulting in a specificity index of 1.00. The DIAL-R was found to correlate moderately with the WPPSI-R (r = .66, p < .01). Correlations between the WPPSI-R and DIAL-R subscales were moderate and significant at the p < .01 level (range r = .37 to r = .69). The DIAL-R produced significantly higher total group scores than the WPPSI-R. Within measure analyses for mean differences for race, gender, age, and risk factors were not significant. Overall, results support the use of the DIAL-R as an accurate and valid screening instrument for preschool children. However, preschool programs should carefully investigate the ramifications of selected cutoffs. If overreferral is the goal, then the
recommended cutoff by the DIAL-R authors of 1.5 SD from the Total score may not be appropriate, given that it failed to identify six children in need of services.
Introduction

In the past twenty years there has been a major thrust to identify preschoolers who are at-risk for academic difficulty and in need of early intervention. This thrust has resulted from research supporting early intervention and laws mandating the provision of early intervention services. Preschool programs, as cost effectively and accurately as possible, must identify children in need of services. Many programs utilize screening instruments to identify children in need of educational remediation. Finding psychometrically sound assessment instruments to identify children in need of services is one problem that has been identified. One aspect of psychometric quality, validity, has been cited as one of the most essential aspects of any assessment measure (Wechsler, 1989). An instrument is valid if it measures what it proposes to measure (Ghiselli, Campbell, & Zedeck, 1981). This researcher proposes to investigate the validity of a commonly used preschool screener, the Developmental Indicators for the Assessment of Learning-Revised (DIAL-R; Mardell-Czudnowski & Goldenberg, 1990).

Early Intervention

The research and the laws supporting early intervention are relatively new and have dramatically impacted the world of education. The following section will provide a brief overview of the legislation that has impacted early intervention services and the efficacy of early intervention.
In the past, research has been focused on diagnosis and treatment of disabled students. However, focus has now shifted from treatment alone to prevention and early intervention (Vacc, Vacc, & Fogleman, 1987). The literature supports the theory that the earlier interventions are begun, the more positive the impact for the child and the lower the subsequent educational cost (Nuttall, Romero, & Kalesnik, 1992; Woods, 1981). According to Woods (1981), if intervention services for a special needs child begins at birth, the cost of serving that child until 18 years of age is $37,273. However, if services do not begin until the child is six, the total cost increases to between $46,816 and $53,340. Therefore, if children with special needs receive early intervention, their prognosis is better and the cost of their education decreases (Nuttall et al., 1992; Woods, 1981). These are both desirable outcomes.

Not only have professionals seen the need for early intervention but the government has seen the need as well. The first federal legislation written exclusively for disabled preschoolers was passed in 1968 (P.L. 90-538). That law provided funds for new approaches to early intervention for children from birth through age eight. In 1972, Head Start programs began setting aside at least 10% of their enrollment capacity for children with disabilities. Building on these laws, Public Law 94-142, the Education for All Handicapped Children Act of 1975, required that if children three to five years old were already being served in that state, all disabled three to five year old children were also to be served. However, this law did not sufficiently meet the needs of preschoolers with disabilities. By the school year 1980-81, only sixteen states required that all three to five year old children with special needs receive services. A
more extensive law designed to better meet the needs of children with disabilities, Public Law 99-457, was passed in 1986. This law has two components. The first component mandates that all states provide children three to five with disabilities a free and appropriate education by the year 1991. The second component provides incentives to states who implement services for children birth through two who experience developmental delays or have an established risk (Heward & Orlansky, 1992).

Much of the knowledge base supporting early intervention has come from findings from government programs. In a review of two articles on the accomplishments of Project Head Start, Zigler (1985) noted that although these programs have shortcomings, they continue to evolve and improve. He further stated that research has repeatedly supported that individualized, multi-resource, family-oriented programs, such as Project Head Start, are exemplary in advancements made in serving preschool children. A study on the efficacy of Model Preschool Program (MPP), which was funded through the Handicapped Children's Early Education Assistance Act, reported positive results from the original program and later programs (Fewell & Oelwein, 1991). The study results revealed significant developmental gains in fine motor skills, communication skills, cognition, and adaptive behavior of children with disabilities. The authors felt that these results supported the foundation upon which P.L. 99-457 was built and the growing trend towards early intervention (Fewell & Oelwein, 1991).

Further evidence supporting early intervention comes from the grant supported Project CARE. The results of this project support that
Early intervention in the form of educational day care plus family support groups resulted in significant cognitive gains after six months (Wasik, Ramey, Bryant, & Sparling, 1990). In addition, Horacek, Ramey, Campbell, Hoffmann, and Fletcher (1987) reported that educational interventions reduced the rate of grade retention and increased achievement scores of children identified at birth as at-risk for school failure. This reviewed research supports the conclusion that early intervention and provisions in the laws that mandate early intervention are necessary to meet the educational needs of children.

Assessment

Assessment of preschool children encompasses many areas. In this section, the distinction between measures used to assess preschool children will be discussed, with the definitions of each provided. The technical adequacy expected of assessment measures will be presented, along with the technical adequacy of measures in the current study. Psychometric aspects relevant to preschool screening instruments will be presented, along with methods of researching these qualities. Lastly, demographic variables will be discussed in relation to how they influence performance on assessment measures.

Preschool Measures. Before at-risk children can receive services through federally funded programs, they must be identified. Many children have obvious disabilities while the disabilities of others are not as easily identified. It is the latter children who most often fail to obtain the benefits of preschool services. Although physicians and agencies refer many children with more severe or obvious disabilities, mass screenings are most effective in identifying those children with milder
disabilities (Jacob, Snider, & Wilson, 1988). The following section will provide a definition and purpose of screening instruments, issues in the assessment of preschool children, and technical adequacy of measures.

The assessment process is designed to identify children with varying degrees of disabilities. This process utilizes instruments on a continuum from screening to diagnostic measures; it is based upon measures that assess a broad range of skills (screening instruments) to measures that thoroughly assess one specific skill (diagnostic instruments). The screening process is the first step in identifying children in need of services. Screening instruments are defined as a brief and cost-efficient assessment procedure. They are used to identify children who may be at-risk for a disability and who may need further evaluation. Screening measures also identify children in need of early intervention programs, or children who require classroom modifications (Barnes, 1982; Meisels, 1985 & 1987; Peterson, 1987). Screening instruments sample a broad range of skills utilizing fewer questions than a diagnostic measure. Language, motor functioning, cognitive ability, and socialization are areas typically assessed by screening measures.

The technical adequacy of screening measures is below that expected of diagnostic measures. Screening instruments sample skills but do not provide a thorough analysis of specific skill strengths and weaknesses. The recommended tendency is for screening instruments to over identify children who may be in need of services. Those who specialize in this area feel that underreferral is more harmful than overreferral (Reynolds & Clark, 1983; Salvia & Ysseldyke, 1981). Basically, screening measures are not appropriate for use in diagnosing
a disability. The design of screening measures produces lower reliability and validity than the design and use of diagnostic measures.

Diagnostic measures are in depth assessment instruments that thoroughly assess one defined area and are used in the process of diagnosing a disability. The purpose of a diagnostic measure requires that the instrument be comprehensive in the specific area(s) assessed and have adequate psychometric qualities. Those children who are identified during the screening process as performing significantly below their age peers may be referred for a diagnostic evaluation. A team of professionals will then make a diagnosis based upon the results of this intense evaluation. If a child performs below the criteria on a preschool screening instrument, a referral may be made to the appropriate professional(s) to determine if there is a developmental deficit. This criteria, or cutoff score, is generally determined by each individual preschool program through the use of options recommended by test publishers.

**Technical Adequacy.** Identifying psychometrically sound measures for assessing preschool children continues to be difficult. Standards for acceptable technical adequacy has been set forth for diagnostic measures and screening instruments (American Psychological Association, 1985). Test developers strive to meet these standards when constructing a new measure, and users look for it when selecting an instrument. Bracken (1987), in a review of several assessment instruments utilized with preschool aged children, recommended specific levels of technical adequacy for these instruments. He reviewed ten preschool measures utilizing his
recommended criteria. Bracken (1987) concluded that each measure had noted strengths and weaknesses; and based upon these, practitioners should carefully select the instrument that best meets the needs of preschool children. Chiefly, the quality of the validity data should be scrutinized and each measure's reliability for the assessed population should be investigated prior to selecting a preschool measure (Bracken, 1987).

Bracken (1987) argues that many preschool assessment measures lack the psychometric qualities demanded of instruments used with older children. The lower psychometric quality compounds the difficulties in assessment due to intrinsic characteristics of preschool children. Characteristics of preschool children are detailed in a subsequent section. Bracken's conclusions regarding invalid and unreliable preschool measures were supported by Meisels (1987). Because of lower psychometric qualities, scores from preschool assessments are frequently used very conservatively as a measure of current functioning rather than a long term predictor of ability (Berk, 1991). These problems with technical adequacy are more predominant with screening instruments than with diagnostic measures at this age.

Studies have established the relationship between intelligence and academic success (Berk, 1991; Sattler, 1990). Children with higher measured IQs tend to get better grades and stay in school longer than those students with lower IQs (Berk, 1991). Therefore, the extent to which a screening instrument can accurately predict intellectual potential will be an important consideration for its use, given IQ is correlated with academic success. The main purpose of a preschool screening measure
is to identify those children who will be at-risk for academic difficulty. Therefore, it stands to reason that screening results may also be correlated with later academic success. This correlation supports the use of preschool assessment as an indicator of children who are at-risk for academic difficulty. Mardell-Czudnowski & Goldenberg (1990) state, "Although the DIAL-R is not intended to be a test of intelligence, it provides a gross estimate of the level of development of intellectual skills needed to succeed in kindergarten and first grade" (p. 67).

Screening measures, which are the type of assessment most often used, are by design less accurate than diagnostic measures. Screening instruments are designed to be brief measures. They are used to assess a broad range of skills and have lower reliability and validity. This design creates two major concerns: 1) not identifying a child as at-risk and 2) identifying a child as at-risk when in fact he/she is not (Miller & Sprong, 1986). To minimize these types of errors one should select and use as accurate a screening measure as possible. Therefore, test users should select and use the instruments with the best psychometric properties. In addition, researchers and test developers should strive to provide this information. The present investigation represents an attempt to provide further information regarding the technical adequacy of one screening instrument, the DIAL-R.

When selecting an instrument, test users look critically at the validity of a measure. Validity studies provide support for the intended use of a measure and provide an indication of the inferences that can be made from the test results (Salvia & Ysseldyke, 1981). The accuracy, or under and overreferral, is an important factor in the selection of a
screening instrument. This accuracy is typically investigated by looking at the classification agreement between a selected screening measure and a criterion measure. The accuracy with which the screening instrument identifies children in need of services is generally referred to as sensitivity. Specificity refers to the accuracy in which the measure detects children not in need of services. By determining the overall classification agreement (correlation), sensitivity and specificity between measures one can make judgments about the validity of these measures. Such information may also provide further data for selecting the cutoff score that will meet the needs of the test user, whether it be a tendency toward over or underreferral.

**Instruments.** The Developmental Indicators for the Assessment of Learning-Revised (DIAL-R; Mardell-Czudnowski & Goldenberg, 1990) provides a measure of early motor, conceptual, and language development for children two to six years of age. The DIAL-R was originally published in 1983 and was revised in 1990. This measure is considered to be one of the best developmental screening measures available (Miller & Sprong, 1986).

Support for the DIAL-R is based upon results from the studies of the previous edition, the DIAL. The authors make this justification in that the DIAL and DIAL-R show significant correlations in all areas at the $p<.05$ significance level. The authors also supported the use of the DIAL technical data for the DIAL-R in the fact that 21 of the 24 items on the DIAL-R are identical or highly similar to the items on the DIAL (Mardell-Czudnowski & Goldenberg, 1990).
Although the authors feel there are no differences, others have questions regarding the reliability and validity of the revised version (Suen, Mardell-Czudnowski, & Goldenberg, 1989). Classification agreement is an important facet of validity for preschool screening instruments, given the main use of these measures is classifying the presence or absence of academic risk variables. Studies investigating the classification consistency of this DIAL-R produced mixed results. A study reported in the manual using the Stanford-Binet as the criterion indicated that the DIAL-R was highly reliable and valid for this purpose (Mardell-Czudnowski & Goldenberg, 1990) and was further supported in a study using the reanalyzed norms (Suen et al., 1989). The squared-error loss generalizability approach was used and indicated that the DIAL-R’s classification reliability supported the measure’s use as a preschool screening measure. All reliabilities for the total scale were .80 or above and decision reliability was p>.95 (Suen et al., 1989). However, studies using other screening instruments, achievement measures, and pupil status as criterion resulted in moderate classification agreement, with a pattern of high specificity (.92 to .95) and low sensitivity (.28 to .57) (Hall & Barnett, 1991; Jacob et al., 1988). This limited research has yet to produce conclusive results, and no studies were found that dealt specifically with the DIAL-R in relation to the criterion measure selected for this study - the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R; Wechsler, 1989).

Assessing the validity of a measure is accomplished through various methods, including examining the content of an instrument and comparing the measure to a valid criterion measure. Comparing a
measure to another measure can assist one in determining the extent to which the two measures assess similar or dissimilar constructs and the accuracy of classification. A diagnostic measure frequently used in the classification process with the preschool population is the WPPSI-R (Wechsler, 1989). The WPPSI-R is considered to be one of the most psychometrically sound instruments available for use with children this age (Bracken, 1987; Kamphaus, 1993).

The Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R; Wechsler, 1989) is an individually administered instrument for assessing the intellectual functioning of children aged three through seven years. It consists of a Verbal and a Performance scale which together yield a Full Scale IQ (FSIQ). The Verbal scale yields a Verbal IQ (VIQ) and the Performance scale a Performance IQ (PIQ). The instrument is currently in its second edition since its first publication in 1967. The WPPSI's technical adequacy has been studied in depth. This trend has continued with substantial studies related to the WPPSI-R. Studies have consistently demonstrated that the WPPSI-R meets or exceeds all validity and reliability criteria and is considered one of the most sound preschool intellectual measures (Kamphaus, 1993). Because of excellent psychometric qualities, it was selected as the criterion measure.

**Demographic Variables.** It is important for screening instruments to assess children equivalently and without bias. However, demographic variables such as race, gender, and age have long been known to influence the results on various measures for various population subgroups. Research in this area is controversial and continues to be
widely studied. The effects demographic variables have on the screening of preschool children is a specific issue related to this study.

There are many things to consider when assessing preschool children. Immense cognitive growth occurs during the preschool years including increased attention, increased memory capacity, and the beginning of an understanding of symbolization. These changes allow for improved learning (Nuttall et al., 1992). However, during the preschool years a range of developmental and behavioral patterns are considered normal (Nuttall et al., 1992). Some children tend to develop in a steady pattern, whereas other children may have little change over periods followed by tremendous spurts (Berk, 1991). Thus, there is a wide range that is considered normal, making it even more difficult to define what is abnormal in the preschool years. These differences may be attributed to any number of factors, among which are the child's background and level of maturity, or development. Factors which are intrinsic to the child contribute to the difficulties in accurately assessing and identifying preschool children with disabilities.

Although the nature of developmental patterns contributes to the difficulties in obtaining high levels of psychometric properties for preschool children, research shows that the measurement of intellectual ability shows greater stability with each passing year until approximately seven years of age (Berk, 1991). In addition, Sattler (1990) indicates that "IQs obtained by the age of five were found to correlate highly with adult IQs (.50 or higher)" (p. 73).

There are other factors that may contribute to difficulties in preschool assessment and assessment of individuals of all ages. These
include performance discrepancies on assessment measures associated with gender, race, and risk factors, such as socioeconomic status (SES). These factors have long been of concern to professionals. Researchers who study child development have established that girls and boys develop at different rates. Overall, boys tend to develop more slowly than girls. Research has shown that girls at the age of five are developmentally one year ahead of their male counterparts (Sherman, 1978). Sherman (1978) was quick to point out that whether or not these differences are similar for cognitive development is still unresolved. Nonetheless, the slower development of boys may produce lower mean scores than girls on developmental measures. The slower development of boys is especially relevant since many measures at the preschool ages are considered to be measures of overall development rather than of cognitive ability.

Gender associated differences have also been noted to significantly affect the performance on assessment measures. Girls have been shown to perform significantly better on verbal related measures and boys significantly higher on visual-spatial tasks (Maccoby & Jacklin, 1974; Schellinger, Beer, & Beer, 1992; Sherman, 1978; Vacc et al., 1987). The majority of these differences have been associated with school-age children (Berk, 1994; Sherman, 1978). However, some differences have been found with preschool age children. Kindergarten aged girls were found to perform better than kindergarten aged boys on a measure of fine motor ability (Sattler, 1990). Sattler also noted that on the Wechsler Preschool and Primary Scale of Intelligence, boys scored significantly higher on the Mazes subtest, whereas girls performed
significantly higher than boys on the Animal House, Geometric Design, Block Design, and Sentences subtests. In contrast, on the WPPSI-R FSIQ no significant mean differences were found between boys and girls (mean age 63.3 months) (Karr, Carvajal, Elser, & Bays, 1993). It should be noted that by comparing same age boys, who are developmentally slower, to same age girls may produce results biased against males, especially at the preschool ages (Sherman, 1978). Sherman (1978) advocates that results from studies supporting sex related differences may show differences (e.g. girls perform better than boys) when in fact it is a developmental difference and not an ability difference. It is important to note that when gender differences are found, they are small and account for no more than five percent of individual differences among children (Berk, 1994). Further research on gender differences in the preschool years may provide a clearer picture of the onset of these performance differences.

The differential performance of racial groups on various assessment instruments has been well documented. The investigation into racial differences is a highly complex and controversial issue. Although researchers have consistently found significant race differences on intellectual measures, the reasons for these differences have yet to be established. It is important to note that these differences are not apparent until between the ages of two and three, but continue through adulthood (Berk, 1994). Whether the difference has a genetic, environmental, or instrument based is still unresolved and continues to be a hotly debated topic. Jensen (1980) feels that racial group differences on intelligence measures are due to heredity. When investigating the difference
between blacks and whites matched on SES, Jensen (1980) found a 12 point difference in favor of whites, with race accounting for 14 percent of the total variance. Other theorists, such as Sandra Scarr (1981), feel that environmental influences account for differences in ability. She cites studies where black children adopted by white families scored above the mean of white children, suggesting that environment does impact measured intelligence.

Recent studies continue to illustrate that blacks, as a group, score significantly lower than whites on assessment measures (Bracken, Sabers, & Insko, 1987; Ittenbach & Harrison, 1990; Sameroff, Seifer, Baldwin & Baldwin, 1993; Vacc et al., 1987). Bracken et al. (1987) support Jensen's theory, that this difference is not test bias but differences in the performance level of the ethnic groups. They also cited several studies which demonstrate that this differences increases in proportion to the child's age.

Although group differences are consistent at the older ages, when looking specifically at the preschool range, the results are varied. Thorndike, Hagen, and Sattler (1986) found a 12 point difference between black children and white children, ages two to six, on the Stanford-Binet IV. However, a study of four to six year old children on the Kaufman Assessment Battery for Children and Stanford-Binet (Form L-M and IV) indicated group differences of no more than four points when SES was controlled (Krohn & Lamp, 1989; Lamp & Krohn, 1990; Vincent, 1991). These results suggest that black/white differences become more apparent at school age. Thus, it is important to study these differences at the preschool level, the time when the onset of the differences appears.
Low SES as a risk factor in the intellectual development of children is a basic premise upon which P.L. 99-457 was built. Socioeconomic status has been shown to strongly relate to intellectual development, with low SES children scoring significantly lower than higher SES children (Largo, Pfister, Molinari, & Kundu, 1989; Mehryar, Tashakkori, Yousefi, & Khajavi, 1987). SES has also been shown useful in the prediction of performance on assessment measures (Vacc et al., 1987). This difference was also found on the WPPSI and is one of the most widely supported factors affecting scores in intelligence measures (Sattler, 1990). A families' socioeconomic status accounts for approximately 11% of the variance in a child's measured intelligence (Sattler, 1990).

Although the relationship between SES and children's test performance has been consistently found and is a firmly established group difference, theories related to the cause vary. Once again, the theories of genetic differences and environmental disadvantage surface (Scarr, 1981; Deutsch, Katz, & Jensen, 1968; Jensen, 1973 & 1980).

It should be noted that demographic variables do not occur in isolation, and each variable impacts the others. In fact, 44% of black children live in poverty, while the national average is 22% (Berk, 1994). This difference appears overwhelming, but Jensen (1980) argues that the combined affect of race and SES accounts for only 22% of the total variance of IQ. Berk (1994) states that even with this small amount of variance, the differences are large enough that they should not be disregarded.
Rationale

The following section will provide support for the current study. First, the need for additional research with the DIAL-R will be described. Then a rationale for the selection of the WPPSI-R as the criterion measure is given. Support for the investigation of differences based on demographic variables is provided. Last, the contributions of this study will be briefly described.

The DIAL was developed with greater attention to technical and psychometric adequacy than most screening instruments. (Jacob et al., 1988). Attempts to improve and update the measure have resulted in the DIAL-R with its reanalyzed norms. In 1983, the test items were updated and revised; then in 1990, the norms were reanalyzed to create an accurate match between the norm sample and the population. Independent research has yet to fully support this current edition.

The DIAL-R is one of the most commonly used instruments for preschool screening (Bracken,1987). The wide use of the DIAL-R creates the need for further information regarding it's accuracy as a preschool screening measure, thereby allowing programs to make better informed choices when selecting a screening instrument. Many three year old preschool programs enroll only children that are identified as at-risk according to federal legislation. It is imperative that the measure utilized be as accurate as possible in classifying children at-risk for academic difficulty so that children in need of services will be served. However, few studies investigating classification validity, an important aspect of validity for preschool screening instruments, were identified. Much of the research conducted on the DIAL-R occurred prior to the 1990
reanalyzed norms. Therefore, studies of the validity of the DIAL-R are greatly needed to support its continued use. The present research will help determine the accuracy with which the DIAL-R identifies children who are at-risk and in need of special services.

The accuracy of a preschool screening measure can be determined by comparing it to a criterion measure. The criterion measure needs to be reliable and valid to produce the most accurate data. Diagnostic intellectual measures are the most reliable and valid measures available for assessing preschool children (Reynolds & Clark, 1983; Salvia & Ysseldyke, 1981). Therefore, one would predict that the more accurately a screening instrument predicts intellectual ability, the more valid the screening instrument (concurrent validity).

The criterion measure selected for this study is the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R). The WPPSI-R was recently revised and has never been compared to the DIAL-R. The current edition of the WPPSI extended the age range and renormed the 20 year old test. The WPPSI-R reported several validity and reliability studies supporting its use. Further research, since its publication, has also provided support for its psychometric qualities. As a result, the WPPSI-R is considered one of the most psychometrically sound preschool intellectual measures (Kamphaus, 1993).

The screening measure should differentiate between students who have identified risk factors, such as developmental delay, and those who do not. It should also identify other students who have academic difficulty or are in need of educational remediation. Those students identified as at-risk on the DIAL-R should perform significantly below
normal students on the criterion measure -- the WPPSI-R. Research has also supported that girls perform higher than boys, whites higher than blacks, and higher SES better than lower SES on assessment measures at school age, with varied results at the preschool age (Bracken et al., 1987; Ittenbach & Harrison, 1990; Mehryar et al., 1987; Sameroff et al., 1993; Schellinger et al., 1992; Vacc et al., 1987). Thus, the researcher will investigate how demographic variables affect results on the DIAL-R and the WPPSI-R at the preschool age. The study results will allow for demographic effects on the DIAL-R to be compared to other measures.

Results from this study will also provide more data to assist in determining the appropriate selection of a screening measure. Correlations and classification agreement results will allow comparisons between screening measures. This study will also contain additional psychometric information to assist in determining the appropriate or continued use of the DIAL-R through the classification agreement with an intellectual measure for an at-risk population. The study results will contribute to the body of information available for decision making.

Research in the area of preschool assessment is something psychologist have desired, preschool program implementors have requested, and the federal law has demanded through requiring accurate identification of children at-risk.

Purpose

The present research is designed to examine the validity of a commonly used preschool screening measure, the Developmental Indicators for the Assessment of Learning-Revised (DIAL-R). Performance on the DIAL-R will be compared to performance on the
Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R), a diagnostic intelligence test. This comparison will provide data on the agreement rate between the DIAL-R and an intellectual measure in identifying children in need of services. It also provides information regarding similarity and differences of the constructs assessed by these two measures. Such a comparison can give an indication as to how well this screening instrument predicts performance on an intelligence measure for a given population.

The present investigator will examine two components of validity. First, the classification agreement between the two measures will be determined. Then, the correlation between the DIAL-R and the WPPSI-R will be examined. Correlation refers to the extent which the scores on one measure (WPPSI-R) can be predicted from the scores on another measure (DIAL-R) (Ghiselli, Campbell, & Zedeck, 1981). Classification agreement determines how accurately the DIAL-R is identifying children in need of specialized services using the WPPSI-R as the selected criterion measure (Mardell-Czudnowski & Goldenberg, 1990). Mean group differences within each measure for various demographic effects will also be investigated. The specific research questions are as follows:

1) What is the extent of classification agreement between the DIAL-R and WPPSI-R for an identified at-risk population?
2) What is the relationship between the Full Scale IQ of the WPPSI-R and the Total score of the DIAL-R? What are the relationships between like subscales of each measure?
3) Is there a mean group difference between the DIAL-R Total and the WPPSI-R Full Scale IQ?
4) Are there significant group total score differences based upon demographic variables for the DIAL-R Total score? Are there significant group differences based upon demographic variables for the WPPSI-R Full Scale IQ?

The following are predicted for each research question:

Hypothesis One: There will be statistically significant positive classification agreement between the DIAL-R and the WPPSI-R as determined by the Chi-Square statistic.

Hypothesis Two: There will be a statistically significant positive correlation between the WPPSI-R Full Scale IQ and the DIAL-R Total score. The DIAL-R Concepts and Language scores will be significantly correlated with the WPPSI-R Verbal Scale. The DIAL-R Motor scale will be significantly correlated with the WPPSI-R Performance Scale. There will be statistically significant positive correlations between the DIAL-R Concepts and Language scores and the WPPSI-R Verbal IQ.

Hypothesis Three: If significant differences are found between the group means of the DIAL-R Total Score and the WPPSI-R FSIQ, the DIAL-R will result in the lower score.

Hypothesis Four: Significant mean differences based upon demographic variables will be present for at-risk versus not at-risk subjects, speech/language versus non speech/language subjects, and developmentally delayed versus not developmentally delayed subjects within the DIAL-R and within the WPPSI-R, with subjects identified as at-risk, speech/language, and developmentally delayed having the significantly lower scores. White subjects will
have significantly higher scores than black/minority subjects. Females with have significantly higher scores than males. Five year old subjects will have significantly higher scores than four year olds.

Method

Subjects

The sample included 54 four and five year old children who entered kindergarten in the fall of 1993. The children ranged in age from 55 to 68 months (SD=3.4). The majority of the children in the program were identified as at-risk based upon a speech/language deficit, a developmental delay, or eligibility for free lunch, with a small portion having no identified risk factor. Table 1 provides descriptive statistics for the sample. The children were enrolled in one of two federally funded preschool programs in western Kentucky during the 1992-93 school year. Children were identified as eligible based on at least one of two criteria. First children could be identified as 'at-risk' as defined by the federal definition, which is eligibility for the federally funded free lunch program. In order to qualify for the free lunch program the family income must be at poverty level, which was $9,061 for a family of one in 1993-94, with an additional $3,198 for each family member added. The remaining criterion for eligibility is through identification of a developmental delay according to Kentucky Administrative Regulations. Children can be identified as developmentally delayed in the areas of language, physical, cognitive, social-emotional, or self-help/adaptive behavior.
Instruments

**DIAL-R.** The Developmental Indicators for the Assessment of Learning-Revised (DIAL-R; Mardell-Czudnowski & Goldenberg, 1990) provides a measure of early motor, conceptual, and language development for children two to six years of age. As reported previously, this measure is considered to be one of the best developmental screening measures available (Miller & Sprong, 1986).

The reliability of the DIAL-R was supported through test-retest, internal consistency, standard error of measurement, and interrater reliability studies. Test-retest reliability scores were: .76 - motor, .90 - concepts, .77 - language, .87 - total. Cronbach's coefficient alpha was utilized to determine the internal consistency of the DIAL-R, with median coefficient ranging from a low of .70 (motor) to .86 (total). Median standard error of measurements were from 2.4 (motor) to 4.6 (total). For interrater reliability, the authors of the DIAL-R based their results upon the results from the DIAL, which indicated an 81 to 99 % agreement (Mardell-Czudnowski & Goldenberg, 1990). These results indicated that the DIAL-R meets psychometric standards for screening measures and is a reliable instrument.

Validity studies supported the DIAL-R and its use as a developmental screening measure. Content validity was supported through the DIAL's close relationship to the DIAL-R. Tryout testing and content analysis for the DIAL-R also supported its content validity. Construct validity was supported in that as a child's age increases, the child's total score will increase ($r=.98$). Correlations between the DIAL-R and other tests were presented by the test developers. However, it was
noted that caution should be exercised in interpreting these studies, in that they were based upon norms prior to the 1990 reanalyzed norms (Mardell-Czudnowski & Goldenberg, 1990). Significant correlations were also reported for the DIAL-R and the K-ABC. Of greatest significance \( (p<.001) \) were the K-ABC Mental Processing Composite as compared to the DIAL-R total, motor and concept scales and the K-ABC Sequential as compared to the DIAL-R total and motor scales.

In studies investigating predictive validity, the DIAL-R was found to be significantly correlated with the kindergarten performance, teacher ratings, the Metropolitan Readiness Test, Clymer-Barrett Readiness Test, and the Stanford Reading Test (Mardell-Czudnowski & Goldenberg, 1990). It should be noted once again that these predictive validity studies were completed using the norms prior to the 1990 reanalysis.

A concurrent validity study was completed using the new norms and the Stanford-Binet, Form L-M. These measures significantly correlated at .68 \( (p<.01) \) (Mardell-Czudnowski & Goldenberg, 1990). Classification agreement, as previously indicated, is not conclusive with DIAL-R authors providing support for the measure and other studies showing low agreement and sensitivity (Hall & Barnett, 1991; Jacob et al., 1988; Mardell-Czudnowski & Goldenberg, 1990; Suen et al., 1989).

The DIAL-R's psychometric qualities were compared with three other screening measures utilizing standards from the APA *Standards for Educational and Psychological Tests* (Miller & Sprong, 1986). The DIAL-R met many of the standards set forth by APA. It was noted that none of the screening instruments reviewed met all the criteria, with the DIAL-R meeting the most, either fully or partially (Miller & Sprong, 1986). The
DIAL-R surpassed all other measures reviewed in that it was the most psychometrically sound screening measure, having met the majority of the test standards required.

**WPPSI-R.** The Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R; Wechsler, 1989) is an individually administered instrument for assessing the intellectual functioning of children aged three through seven years. It consists of a Verbal and a Performance scale which together yield a Full Scale IQ. The instrument is currently in its second edition since its first publication in 1967. The WPPSI’s technical adequacy has been studied in depth. This trend has continued with substantial studies related to the WPPSI-R.

Reliability studies using the WPPSI-R are extensive and are generally supportive of the WPPSI-R as a reliable measure. Split-half reliability averaged across the ages was .80 or better, except for Object Assembly, Geometric Design, and Mazes which were .63, .79, and .77, respectively. The Performance IQ, Verbal IQ, and Full Scale IQ were .92, .95, .96, respectively (Wechsler, 1989). Standard error of measurement, another index of reliability, for the PIQ, VIQ, and FSIQ ranged from 2.81 to 4.98. Interscorer reliability was found to range from .88 to .96 on the most subjectively scored subtests. A three to seven week test-retest study revealed a coefficient of .88 for the PIQ, .90 for the VIQ, and .91 for the FSIQ (Wechsler, 1989). These findings support the WPPSI-R as a reliable measure.

Studies of validity support the use of the WPPSI-R as an accurate measure of cognitive ability. A study by Faust and Hollingsworth (1991), compared the McCarthy Scales of Children's Abilities (MSCA), the
Peabody Picture Vocabulary Test-Revised (PPVT-R), and the WPPSI-R. The MSCA is a commonly used measure of cognitive ability in the preschool years. The PPVT-R measures receptive vocabulary and is often used as an intellectual screening measure. The examiners counterbalanced administration of the tests and found no difference in test scores based upon order of measure given. They also found no gender differences. Low correlations, in the .30 range, were found between the WPPSI-R and the PPVT-R. The WPPSI-R (FSIQ) and the MSCA (GCI) produced a stronger relationship (.67), as did the Verbal scales of each with correlations of .65. The stronger relationship between the WPPSI-R and the MSCA, as compared to that between the PPVT-R and the WPPSI-R was expected. The MSCA is an intellectual measure similar to the WPPSI-R, whereas the PPVT-R is a screening instrument. Higher correlations would be predicted between the DIAL-R and WPPSI-R, given that the design of the DIAL-R is more similar to that of the WPPSI-R. The breadth of the DIAL-R (concepts, motor, and language) is more closely related to the WPPSI-R measure of verbal and performance tasks than that of the PPVT-R.

Validity of the WPPSI-R is further supported in the analysis of the intercorrelation with other measures. A study on the intercorrelation of the WPPSI-R subtests suggests a common component that each subtest measures; however, there is still a unique aspect to each subtest. Factor analysis supported the theory of interpretation of the performance and verbal abilities as separate entities (Gyurke, Stone, & Beyer, 1990; LoBello & Gulgoz, 1991; Wechsler, 1989). The WPPSI-R was highly correlated with the WPPSI (.87), the Wechsler Intelligence Scale for
Children-Revised (.85), the Stanford-Binet Intelligence Scale, Fourth Edition (.74), and the McCarthy Scales of Children's Abilities (.81). The Kaufman Assessment Battery for Children showed moderate correlations at .55 (Kamphaus, 1993; Wechsler, 1989). High correlations with these measures strongly support the construct validity of WPPSI-R, given that these measures are the most widely accepted intellectual measures.

Procedure

Permission forms were sent home with the children in the preschool program for their parents to sign and return to the classroom teacher (see Appendix A). The WPPSI-R and the DIAL-R were administered by examiners who had graduate training in individual assessment and psychometrics. Each examiner was trained to administer the assessment instruments according to standardized procedures indicated by the manual of each instrument.

The examiners spent time in each classroom so that the children could become familiar with the examiners. It was explained to the class that some of the children were going to "play" with the examiners at different times. Each child was then tested in the order assigned with no more than one hour of testing per day. The testing occurred so that the DIAL-R and the WPPSI-R were counterbalanced. After each testing session the child received a sticker for participating.

Analysis

The analysis of this study is descriptive in design. To address the first question (What is the extent of classification agreement between the DIAL-R and WPPSI-R?) a frequency distribution table was developed to determine the agreement rates in classifying children into potential
problem and okay/normal categories. Criteria selected for potential problem on the DIAL-R is that recommended by the authors, 1.5 standard deviations below the Total score mean. The WPPSI-R criteria is 78, which is approximately 1.5 standard deviations below the mean. The extent to which the DIAL-R correctly identifies children who have a potential problem is defined as sensitivity (true positives). Specificity refers to the extent to which the DIAL-R correctly identifies those children who do not have potential problems (true negatives). The agreement index is the proportion of agreement between the two measures on children who are okay and have a potential problem (Hall & Barnett, 1991; Mardell-Czudnowski & Goldenberg, 1990). The Chi Square test was used to determine if significant differences existed between observed classification agreement and that expected by chance. To establish the correlation between the WPPSI-R and DIAL-R a Pearson R was conducted. A t-test was conducted to determine if there is a significant mean difference between the DIAL-R Total and the WPPSI-R FSIQ for the total group. The affect of demographic variables for both the DIAL-R Total score and the WPPSI-R FSIQ was investigated through independent t tests. Dunn Bonferroni correction will be used to control the overall error rate to p<.05.

Results

The first hypothesis in the current study involved the extent of classification agreement between the DIAL-R and WPPSI-R. Table 2 illustrates the decision classification of subjects based on results obtained from the two measures. The four cells of the table include 1) children identified as at-risk by the DIAL-R and the WPPSI-R;
2) children identified as at-risk on the WPPSI-R, but not at-risk on the DIAL-R; 3) children identified as not at-risk on the WPPSI-R, but at-risk on the DIAL-R; and 4) children identified as not at-risk on the DIAL-R and WPPSI-R.

The agreement index suggests a high rate of agreement (.89) between the two measures when the cutoff of 1.5 SD for the total score is used. The specificity index (true negatives) indicated perfect agreement (1.00); sensitivity index (true positives) was low (.40). Thus, the DIAL-R appropriately identified four children as at-risk but failed to identify six children identified by the criterion measure. It did not overrefer any children when the total score cutoff of 1.5 SD was used. The phi coefficient was statistically significant at the .00001 level, an indication that agreement on risk status exceeded what would be expected by chance (Chi-Square (1) = 19.01, p<.00001).

The correlation between the DIAL-R Total and WPPSI-R FSIQ, along with the subscales of each, was the second hypothesis addressed in the study. The Pearson correlations of the DIAL-R area and total score with the WPPSI-R IQ's are presented in Table 3. All correlations were moderate and significant at the p< .01 level. The WPPSI-R FSIQ showed the highest correlations with the DIAL-R Concepts (.69) and Total (.66). The lowest correlation was found between the WPPSI-R PIQ and the DIAL-R Concepts, which were significantly correlated (.37) but accounted for the less than 14% of the variance. Moderately strong correlations are noted between the DIAL-R Total and the WPPSI-R PIQ, VIQ, and the FSIQ (r =.50, .65, and .66, respectively). A moderately strong correlation
was also noted between the WPPSI-R Verbal Scale and the DIAL-R Concepts Scale (r = .65).

Whether or not subjects performed equally on the total score of the DIAL-R and the WPPSI-R was examined in the third hypothesis. This hypothesis and hypothesis four were investigated through the use of an independent t tests utilizing the Dunn's Bonferroni correction to control the error rate for the set of analyses at p<.05 critical value. Table 4 provides the mean, standard error of measurement, and standard deviation for the DIAL-R Total and WPPSI-R FSIQ. Results of the t test, illustrated in Table 5, indicated that there was a significant difference between the DIAL-R Total and the WPPSI-R FSIQ, t(53)= 12.56, p<.05. The difference between the DIAL-R and the WPPSI-R ranged from 1 to 36 in favor of the DIAL-R, with only 3 subjects having higher WPPSI-R scores (see Table 6). Approximately 57% of the subjects had a difference of 1 SD or more, with a mean difference of 16.37. Table 7 provides the mean, standard error of measurement, and standard deviation for each demographic variable. The mean differences on the variables ranged from 13.33 (S/L) to 19.81 (Not AR) in favor of the WPPSI-R.

In the last hypothesis the mean differences were examined within each measure based upon demographic variables for the DIAL-R Total and the WPPSI-R FSIQ. The results of t tests, with the Dunn's Bonferroni correction, indicated no significant mean differences for either measure based upon demographic variables. Table 8 provides these results.

Discussion

In this section the results will be examined in relation to several factors. The findings will first be discussed in relation to the hypotheses
generated in this study and in relation to past studies. How these affect
the practical use of the DIAL-R and any theoretical implications will be
investigated next. Then, any limitations that affect the interpretation of the
data will be examined. Last, suggestions for future research in this area
will be presented.

Examined in this study was the classification agreement between
the DIAL-R and the WPPSI-R. Study results indicated a significantly high
agreement rate between the DIAL-R and the WPPSI-R, which supports
Hypothesis One. The DIAL-R manual supports the tendency toward
overreferral versus underreferral through the design of the screening
measure and the recommended 1.5 SD cutoff score (Mardell-
Czudnowski & Goldenberg, 1990). The current study did not detect
overreferral when using an intellectual measure as the criterion. Rather,
it was found that the DIAL-R did not identify six subjects in need of
services, and none of the subjects were overreferred. The finding of
underidentification of children in need of services may be due to the
components on the DIAL-R that are dissimilar to the WPPSI-R construct.
For example, the motor and language components have equivalent
decision weight as the concepts area but are not as highly correlated
with the WPPSI-R. Thus, the fact that the DIAL-R measures a broader
range of skills may create this discrepancy. Therefore, it appears that for
subjects similar to the ones in this study the criteria of 1.5 SD may be too
lenient if the goal is toward overreferral.

These results are different from those found when using the
Stanford-Binet, Form L-M as the criterion (Mardell-Czudnowski &
Goldenberg, 1990). Results from that study suggested that the DIAL-R
overreferred four times as many subjects as it underreferred. However, studies in the WPPSI-R manual indicate that the Stanford-Binet's mean score was significantly higher than the WPPSI-R (seven points), but other studies showed as little as a two point difference (Wechsler, 1989). This difference in results when comparing the WPPSI-R and Stanford-Binet to the DIAL-R may be attributed to this study's restricted sample size, the skewed at-risk population, or differences between the criterion measures themselves.

To further investigate the validity of the DIAL-R for identifying children at-risk for academic difficulty, the sensitivity and specificity were examined. The sensitivity, which is the extent to which the DIAL-R correctly identified children as in need of services, was .40 using the total score cutoff of 1.5 SD for the DIAL-R and a standard score of 78 on the WPPSI-R. Specificity was 1.00. Specificity is an index of the proportion of children the DIAL-R correctly identified as not having a potential problem. These results were similar to those found in the study by Jacob et al. (1988) which found the DIAL-R's sensitivity and specificity to be .43 and .97, respectively, using current pupil status as the criteria. It was noted that in comparison to the Early Screening Inventory (ESI), the DIAL-R had higher specificity and lower sensitivity (Jacob et al., 1988). A similar pattern of high specificity and low sensitivity was found in this study. This consistency gives further support that when using the total score cutoff of 1.5 SD, the DIAL-R tends to underrefer rather than overrefer.

The concurrent validity of the DIAL-R with the WPPSI-R for at-risk preschool children was examined in the second component of this study.
Moderate statistically significant positive correlations were found between the DIAL-R Total and area scores and the WPPSI-R IQs. These results support Hypothesis Two. The correlation with the DIAL-R Total score and the WPPSI-R FSIQ ($r = .66$) is consistent with DIAL-R correlations with the Stanford-Binet ($r = .68$) and slightly higher than that found for the Kaufman Assessment Battery for Children ($r = .55$) (Mardell-Czudnowski & Goldenberg, 1990). These results support the use of the DIAL-R to identify preschool children who have below average level of intellectual skills and may be at risk for academic difficulty. However, it should be noted that the correlation between the DIAL-R and WPPSI-R accounts for only 43% of the variance; therefore, the constructs while similar, also have dissimilar components. For example, the motor component on the DIAL-R has a gross motor component that the WPPSI-R does not have.

In order to provide more practical information the mean score for the DIAL-R and for the WPPSI-R were compared. Although a significant mean difference was noted, its direction was different than hypothesized. It was hypothesized that the WPPSI-R would produce the higher score, when in fact the DIAL-R produced the higher score. The mean difference was more than one full standard deviation from the mean (16.37). This difference may be related to the fact that the instruments were standardized in different years, that the DIAL-R is a less reliable measure, or that they measure different constructs (Bracken, 1988). Therefore, when using the DIAL-R with samples similar to the one used in this study, the clinician can expect scores on the WPPSI-R to be approximately one standard deviation lower than the DIAL-R scores.
The fourth hypothesis was fully rejected. No demographic variables accounted for significant mean differences on either the DIAL-R or the WPPSI-R. Not only were there no gender, race, or age differences, there were no differences for identified risk variable as well. The latter finding is a major concern, given the major use of the DIAL-R is to differentiate between children in need of services and additional assessment and those children in the normal range.

The absence of differences due to demographic variables may be due to several factors. A population of children who have various risk variables may perform similarly on assessment measures, therefore eliminating any apparent demographic differences. The low number of subjects in a restricted sample may have prevented significant differences. Lastly, differences on the demographic variables of gender, race, and SES may not appear at these ages or may not be of significant magnitude to produce statistical significance. Although differences based upon the aforementioned demographic variables are consistently found with school age subjects, they are not as consistent at the preschool age. Some studies on gender and race illustrate that these differences are not consistently found at the preschool age (Karr et al., 1993; Krohn & Lamp, 1989). The current findings support these recent studies. The fact that differences are not found consistently at the preschool level may provide further evidence for when and why these differences appear.

The results of this study provide valuable information for professionals when selecting a preschool measure. The major use of preschool screening instruments is the identification of children in need
of services or further assessment. Given that a screening measure is not used to make a diagnosis, but to identify those children who have potential problems, it is generally recommended that there be a bias toward identifying a child as at-risk. Identifying a normal child as at-risk is judged to be less harmful than not identifying a child who is at-risk (Reynolds & Clark, 1983). Salvia and Ysseldyke (1981) go further to state that the only errors a screening instrument should make are false positives -- identifying normal children as at-risk for a handicap.

The overreferral of children was not found with this sample using the Total score cutoff of 1.5 SD. The results demonstrated that the DIAL-R produced significantly higher scores than a diagnostic measure (WPPSI-R) produced. The significant mean difference is not only important information for the clinician and preschool screening coordinator but it also has strong implications for the sensitivity of the test. As seen in the Jacob et al. study (1988) and in the present study, the higher DIAL-R scores reduce the sensitivity of the test and allow many children in need of services to go undetected.

Although the test appears valid for assessing children at-risk, the DIAL-R suggested cutoff of 1.5 SD from the Total score mean may not be the most appropriate cutoff for samples similar to the ones in the current study, which includes a high number of children identified as at-risk, having developmental delays, or having speech/language deficits. To increase the sensitivity of the DIAL-R an alternative cutoff of 1 SD or potential problem in one area may be used which may identify more children. The professionals must weigh the ramifications of each cutoff and choose the cutoff that will best serve the needs of the program. If the
preference is toward overreferral, the 1 SD cutoff may be most appropriate; however, if the preference is toward underreferral then 1.5 or 2 SD may be most appropriate. Using one low area (motor, concepts, or language) or local norms for the test for the decision to refer was found to be the most effective in identifying children in need of services (Jacob et al., 1988). With overreferral comes the added cost of giving full diagnostic batteries to all children identified during the screening process. However, with underreferral the increased cost comes later in the increased services needed if child is identified later in their school career. It should be reiterated that professionals need to consider the ramifications of each cutoff choice prior to making a selection.

The DIAL-R is used to identify children in need of special services. One way to investigate the DIAL-R's accuracy is to compare the scores of children previously identified as at-risk by a diagnostic measure or other criteria measures to scores of normal children, or those not identified as at-risk for academic difficulty. One would expect that children identified as at-risk would score significantly lower than children not identified as at-risk. However, children previously identified as developmentally delayed, speech/language, and at-risk did not perform statistically different from children not identified in these areas. For example, children identified as developmentally delayed prior to the study did not perform statistically different on the assessment measures from children not identified as developmentally delayed. Although professionals should consider this information, it should also be noted that the majority of the children in the study were identified in one of the three areas. Therefore, when the DIAL-R did not differentiate between
developmentally delayed and non-developmentally delayed it may have been because the latter group had another risk variable (speech/language or SES) that produced scores similar to those of developmentally delayed children. This bias holds true for all the comparisons for risk factors.

These results impact the future development of preschool screening measures. Developers of screening instruments should be sensitive to the needs of those who use preschool screening instruments. The accepted tendency is towards overreferral; therefore, this bias needs to be built into the measure. The fact that the DIAL-R score results in more than 1 SD higher than the WPPSI-R in the present study demonstrates that the bias to over identify is not present for the at-risk population. However, the high risk population is the population which the screening instrument targets. These findings suggest that a change in the interpretation of the DIAL-R may be needed. An alternative method of looking at each individual component for a deficit may offer more information than the Total score. For example, a child may have a language deficit, but normal motor and concept ability may mask this deficit. The DIAL-R is one of the most accepted preschool screening instruments used. The results of this study suggest that screening instruments need further improvements before we can more accurately identify preschool children in need of services. The reliability and validity of screening instruments, especially for at-risk populations, should be fully investigated during their development and prior to consideration for use.
All studies have limitations, and results should be interpreted in light of these limitations. One limitation of the present study involves the population assessed. The majority of the subjects in this study were identified as at-risk due to SES or an identified disability and thereby created difficulty in generalizing these findings to more heterogeneous populations. However, at-risk students are the ones that tend to be of greatest concern for preschool programs. The children in this study are from two preschool programs, restricted to one small geographic area and may not be representative of national programs. The poorer reliability of the DIAL-R as compared with that of the WPPSI-R may explain some of the differences found. For example, the DIAL-R is not as reliable a measure, therefore it does not assess skills as accurately as the WPPSI-R, creating errors in identifying children in need of services. An additional limitation is that these children had been participating in a preschool program designed to meet the needs of at-risk children. This intervention could have significantly increased the students scores on DIAL-R subtests, which are similar to skills targeted in the preschool setting. Lastly, the parents who choose to let their children be involved in the study may be a different population than those who did not want their child to participate.

Future research needs to further explore the DIAL-R's validity as a preschool screening instrument. Specifically, research to determine the accuracy with which the DIAL-R identifies children in need of services should be addressed. It would be important to determine if school age children identified as mild mentally disabled or specific learning disabled were detected as having potential problems during preschool screening.
Such a predictive validity study could be done through a longitudinal study by following children from preschool through their school years or through a retrospective archival study. Another area of future study involves a comparison between the DIAL-R results and results from a comprehensive diagnostic battery. A diagnostic battery is used in the "real world" to determine a disability; therefore, it would be appropriate to assess the classification agreement between the DIAL-R and a diagnostic battery.

The finding of a mean difference between the DIAL-R Total score and the WPPSI-R FSIQ needs to be further investigated. This finding of higher DIAL-R Total scores than WPPSI-R FSIQ for an at-risk population needs to be replicated. Mean differences should also be investigated based upon demographic variables. For example, does the DIAL-R always produce higher scores than the WPPSI-R for males, females, blacks, whites, at-risk, etc. In addition, it would be important to determine if this difference is apparent in a normally distributed population.

It is necessary for a preschool screening instrument to have high sensitivity. Results from the current study and previous studies indicate that when using 1.5 SD as the cutoff criteria the accuracy with which the DIAL-R identifies children in need of services (sensitivity) is low. Therefore, investigating ways to increase sensitivity of the DIAL-R is essential. Comparing the classification agreement between a criterion measure and the DIAL-R cutoff criteria of 1 SD, 1.5 SD, and 2 SD will provide further information for selecting the appropriate criterion to meet the preschool programs need. Specifically, reanalyzing the current data using 1 SD as the cutoff will provide information on increasing the
specificity of the DIAL-R. Additional data from this study indicated that there were no significant differences found within groups based upon demographic variables (i.e., at-risk versus not at-risk) on the DIAL-R or WPPSI-R. However, the majority of the subjects in the current study had one or more risk variables. Therefore, the children identified as not at-risk may have had a developmental delay or other risk factor which may have masked any differences. Investigation of at-risk groups' performance compared to a "normal" group's performance will provide further information related to how well the DIAL-R distinguishes between at-risk and normal children. This comparison was investigated during the standardization of the DIAL-R using the Stanford-Binet as the criterion measure. However, the WPPSI-R and Stanford-Binet produce slightly different scores for this age group making it worthwhile to also investigate group differences using the WPPSI-R. As noted in this study, the population consisted of children who had one or more risk variables. These included established developmental delay, speech/language, and low SES. It would be beneficial to investigate which of these risk factors, or combination of risk factors, best predict identification as potential problem on the DIAL-R. The variance related to cumulative risk factors may offer a better explanation for why children perform in the potential problem range on the DIAL-R than any single risk factor.
Table 1

Subject Demographic Characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>Percent</th>
<th>$\bar{x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Subjects</td>
<td>54</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Age in Months</td>
<td></td>
<td></td>
<td>60.87</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minority</td>
<td>15</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>39</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>20</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>34</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>SES$^a$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At-Risk</td>
<td>38</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Not At-Risk</td>
<td>16</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Identified Disabilities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech/Language</td>
<td>21</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Developmental Delay</td>
<td>6</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

$^a$At-risk is defined as eligibility for the federally funded free lunch program.
Table 1

Decision-Making Validity of the DIAL-R when Compared with the WPPSI-R on 54 Preschool Children

<table>
<thead>
<tr>
<th>WPPSI-R Decisiona</th>
<th>Potential Problem</th>
<th>Okay</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DIAL-R Decisionb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential Problem</td>
</tr>
<tr>
<td>(40% sensitivity)</td>
</tr>
<tr>
<td>Okay</td>
</tr>
<tr>
<td>(100% specificity)</td>
</tr>
</tbody>
</table>

Note. Chi-Square = 19.01; df = 1; \( p < .00001 \). Agreement index = .89.

aPotential Problem on the WPPSI-R corresponds to a score of 78 or less.

bDIAL-R decisions are based on the DIAL-R Total score decision method, using 1.5 standard deviation cutoffs.
Table 3

Correlation between the DIAL-R Area and Total Scores and WPPSI-R IQ Scores for a Sample of 54 Preschool Children

<table>
<thead>
<tr>
<th></th>
<th>Motor</th>
<th>Concepts</th>
<th>Language</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>WPPSI-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>.46*</td>
<td>.65*</td>
<td>.54*</td>
<td>.65*</td>
</tr>
<tr>
<td>Performance</td>
<td>.43*</td>
<td>.54*</td>
<td>.37*</td>
<td>.51*</td>
</tr>
<tr>
<td>Full Scale</td>
<td>.51*</td>
<td>.69*</td>
<td>.52*</td>
<td>.66*</td>
</tr>
</tbody>
</table>

*p < .01
### Table 4

**Mean, Standard Error of Measurement, and Standard Deviation for the DIAL-R and the WPPSI-R**

<table>
<thead>
<tr>
<th></th>
<th>Mean (x)</th>
<th>SEM</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WPPSI-R</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal Scale IQ</td>
<td>88.02</td>
<td>2.28</td>
<td>10.18</td>
</tr>
<tr>
<td>Performance Scale IQ</td>
<td>87.43</td>
<td>3.40</td>
<td>12.02</td>
</tr>
<tr>
<td>Full Scale IQ</td>
<td>86.35</td>
<td>2.08</td>
<td>10.40</td>
</tr>
<tr>
<td><strong>DIAL-R</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motor</td>
<td>98.52</td>
<td>5.99</td>
<td>10.94</td>
</tr>
<tr>
<td>Concepts</td>
<td>101.09</td>
<td>4.98</td>
<td>10.17</td>
</tr>
<tr>
<td>Language</td>
<td>106.65</td>
<td>7.91</td>
<td>14.95</td>
</tr>
<tr>
<td>Total</td>
<td>102.72</td>
<td>4.67</td>
<td>12.48</td>
</tr>
<tr>
<td></td>
<td>DIAL-R Total</td>
<td>WPPSI-R FSIQ</td>
<td>Mean Difference (DIAL-R - WPPSI-R)</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td><strong>x</strong></td>
<td>102.72</td>
<td>86.35</td>
<td>16.37</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>12.48</td>
<td>10.40</td>
<td>9.58</td>
</tr>
<tr>
<td><strong>t-value</strong></td>
<td></td>
<td></td>
<td>12.56</td>
</tr>
<tr>
<td><strong>2-tail Probability</strong></td>
<td></td>
<td></td>
<td>.05</td>
</tr>
<tr>
<td><strong>Eta Squared</strong></td>
<td></td>
<td></td>
<td>.59</td>
</tr>
</tbody>
</table>
Table 6

Frequency of the Difference between the DIAL-R Total and WPPSI-R FSIQ

<table>
<thead>
<tr>
<th>Difference Value</th>
<th>Frequency</th>
<th>Difference Value</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-6.00</td>
<td>1</td>
<td>17.00</td>
<td>5</td>
</tr>
<tr>
<td>-4.00</td>
<td>1</td>
<td>18.00</td>
<td>2</td>
</tr>
<tr>
<td>-3.00</td>
<td>1</td>
<td>19.00</td>
<td>2</td>
</tr>
<tr>
<td>1.00</td>
<td>1</td>
<td>20.00</td>
<td>1</td>
</tr>
<tr>
<td>3.00</td>
<td>1</td>
<td>21.00</td>
<td>5</td>
</tr>
<tr>
<td>5.00</td>
<td>1</td>
<td>22.00</td>
<td>1</td>
</tr>
<tr>
<td>6.00</td>
<td>1</td>
<td>23.00</td>
<td>3</td>
</tr>
<tr>
<td>7.00</td>
<td>2</td>
<td>24.00</td>
<td>1</td>
</tr>
<tr>
<td>8.00</td>
<td>1</td>
<td>25.00</td>
<td>1</td>
</tr>
<tr>
<td>9.00</td>
<td>3</td>
<td>26.00</td>
<td>2</td>
</tr>
<tr>
<td>10.00</td>
<td>1</td>
<td>27.00</td>
<td>3</td>
</tr>
<tr>
<td>11.00</td>
<td>3</td>
<td>29.00</td>
<td>1</td>
</tr>
<tr>
<td>12.00</td>
<td>3</td>
<td>31.00</td>
<td>1</td>
</tr>
<tr>
<td>13.00</td>
<td>1</td>
<td>34.00</td>
<td>2</td>
</tr>
<tr>
<td>14.00</td>
<td>2</td>
<td>36.00</td>
<td>1</td>
</tr>
<tr>
<td>15.00</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7

Mean, Standard Error of Measurement, and Standard Deviation for the DIAL-R and the WPPSI-R Total Scores for Each Demographic Variable

<table>
<thead>
<tr>
<th></th>
<th>DIAL-R</th>
<th>WPPSI-R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>SEM</td>
</tr>
<tr>
<td>Minority</td>
<td>102.33</td>
<td>4.71</td>
</tr>
<tr>
<td>White</td>
<td>102.87</td>
<td>4.73</td>
</tr>
<tr>
<td>Male</td>
<td>104.26</td>
<td>4.23</td>
</tr>
<tr>
<td>Female</td>
<td>100.10</td>
<td>5.34</td>
</tr>
<tr>
<td>AR</td>
<td>100.57</td>
<td>4.71</td>
</tr>
<tr>
<td>Not AR</td>
<td>107.81</td>
<td>4.13</td>
</tr>
<tr>
<td>S/L</td>
<td>95.62</td>
<td>5.73</td>
</tr>
<tr>
<td>Not S/L</td>
<td>107.24</td>
<td>3.26</td>
</tr>
<tr>
<td>DD</td>
<td>91.33</td>
<td>6.56</td>
</tr>
<tr>
<td>Not DD</td>
<td>104.15</td>
<td>4.19</td>
</tr>
</tbody>
</table>

aAR=At-Risk  bS/L=Speech/Language  cDD=Developmental Delay
Table 8

T-test Scores for Mean Differences on the DIAL-R and WPPSI-R Total Scores for each Demographic Variable

<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>DIAL-R t-value</th>
<th>WPPSI-R t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race</td>
<td>.14</td>
<td>1.00</td>
</tr>
<tr>
<td>Gender</td>
<td>1.19</td>
<td>.49</td>
</tr>
<tr>
<td>Age</td>
<td>.52</td>
<td>.47</td>
</tr>
<tr>
<td>SES</td>
<td>2.00</td>
<td>.75</td>
</tr>
<tr>
<td>S/L</td>
<td>3.72</td>
<td>2.39</td>
</tr>
<tr>
<td>DD</td>
<td>2.48</td>
<td>3.00</td>
</tr>
</tbody>
</table>

*Minority versus white *Male versus Female *four year olds versus five year olds *AR versus Not AR *S/L versus Not S/L *DD versus Not DD
Appendix A

Permission forms were created and okayed by the teachers of each preschool program. The permission form was sent home to obtain parental permission and to provide a brief description of the study.
Would you like to find out more about your child's developmental status?

Your child is invited to participate in a study involving tests which will help educators evaluate social skills, motor ability, and skills related to school success (screeners) and ability to learn (intelligence). These activities are designed to be fun, interesting, and entertaining for preschool children. These screeners are similar to other preschool tests which are administered in preschools around the country.

This study is being conducted by Shannon Batchelor and Wendy Simmons, School Psychology graduate students at Western Kentucky University, in cooperation with your child's school. The testing will take place in your child's school during regular class hours. Before the end of the school year feedback will be available to each parent regarding his/her individual child's test results.

We emphasize that your child's participation in this project is entirely voluntary. All information specifically about your child will be kept strictly confidential and will only be seen by the research staff. The child and the school will never be identified by name.

Please respond by March 25. If you have any questions about this study, please call Shannon Batchelor at 726-6156, Dr. Elizabeth Jones at 745-4414, or Wendy Simmons at 781-0763. We will be happy to hear from you.

*Tests to be administered
Denver Developmental Screening Test-Revised II
Early Screening Profiles
Wechsler Preschool and Primary Screening for Intelligence-Revised
Participant Consent Form

Dear Parents:

Your child is invited to participate in a study of children's performance on tests of readiness (screeners) and ability to learn (intelligence). This study is being conducted by Shannon Batchelor and Wendy Simmons, School Psychology graduate students at Western Kentucky University, in cooperation with your child's school. The screeners provide information about your child in areas related to social skills, motor (movement) ability, and skills related to school success. The intelligence test will provide information about your child's ability to learn. These instruments are designed to be fun, interesting, and entertaining for preschool children. The researchers will then see how well each test measures readiness skills in preschool aged children and how well they predict ability to learn. The information gained will help teachers, school counselors, and school psychologists interpret the results of these tests to better meet the needs of students. The screening tests that will be administered are the Denver Developmental Screening Test-Revised II and the Early Screening Profiles. These screeners are similar to the Developmental Indicators for the Assessment of Learning-Revised which will be administered to your child by the school and will be compared to the screeners and cognitive measures administered by the researchers. In addition to the screeners the Wechsler Preschool and Primary Screening of Intelligence-Revised, an instrument that assesses overall learning ability, will be administered.

The testing will take place in your child's school during regular class hours in March and April for approximately seven days. However, each child will only take approximately 20 to 30 minutes for each screener and approximately 1 hour to 1 1/2 hours for the intelligence test. Testing will be spaced out over two or three days for each child. Before the end of the school year feedback will be available to each parent regarding his/her individual child's test results.

We emphasize that your child's participation in this project is entirely voluntary. If you or your child decide not to participate, it will have no negative outcome for you or your child in any way. You are free to withdraw consent and discontinue participation at any time during the testing without any negative consequences. All information specifically about your child will be kept strictly confidential and will only be seen by the research staff. All results will be reported in terms of averages of groups of children and children will never be identified by name.

If you have any questions about this study, please call Dr. Elizabeth Jones at 745-4414, Shannon Batchelor at 726-6156, or Wendy Simmons at 781-0763. We will be happy to hear from you.
Please Return This Page To Your Child's Teacher

We hope that you will allow your child to take part in our study. We promise to do our best to make it a pleasant experience and to schedule our sessions in cooperation with your child's teacher. To indicate your consent, please fill in your child's name, teacher, and sign your name.

Participant Consent Form

Child's Name:_________________________ Teacher: ________________

I have read the information provided about this study, and give my consent for my child to participate in the screening and cognitive assessment and allow the researchers access to the results of the Developmental Indicators for the Assessment of Learning-Revised that are part of my child's school records. I understand that I may withdraw my child from the study at any time without penalty.

Parent's Signature:_________________________ Date:______________

Permission for Release of Test Results

Given the confidential nature of these results it is necessary to obtain your permission before we release the results to the school. It is not necessary to release the test results to the school in order to participate in the study. You as a parent will receive the test results whether or not you give permission to release the results to the school. If released, the results may provide information that will help the school better meet your child's educational needs. You may withdraw permission for release at any point. If you wish to have your child's results placed in his/her school records please sign below.

Parent's Signature:_________________________ Date:______________
Appendix B

The parents were given written reports of their child's assessment results. Times were also scheduled for the parent to receive verbal explanation of the results.
Report to Parents

Thank you for allowing your child to participate in our preschool study. Testing has been completed and the reported results are as follows:

Report for:
Name: ___________________________ Age ______
Parents' Names ___________________________
School: ___________________________

Tests Administered:
1. Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R): The WPPSI-R provides a measure of verbal reasoning ability (Verbal Scale) and nonverbal reasoning ability (Performance Scale), which together yields a Full Scale score. Your child will perform in this range 95 out of 100 times.

<table>
<thead>
<tr>
<th>Score Range</th>
<th>Functioning Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Scale</td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td></td>
</tr>
</tbody>
</table>

2. The Developmental Indicators for the Assessment of Learning-Revised (DIAL-R): The DIAL-R screens development of motor skills, concepts, and language skills.

<table>
<thead>
<tr>
<th>Score Range</th>
<th>Functioning Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIAL-R Total</td>
<td></td>
</tr>
</tbody>
</table>

3. The Early Screening Profiles (ESP): The ESP provides a measure of the child's developmental status in the areas of cognitive/language, motor, and age appropriate behavior necessary for daily functioning.

<table>
<thead>
<tr>
<th>Score Range</th>
<th>Functioning Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESP Total</td>
<td></td>
</tr>
</tbody>
</table>

4. The Denver Developmental Screening Test-II (DENVER II): The DENVER II provides a measure of the child's development in the areas of personal/social, fine motor, language, and gross motor skills.

<table>
<thead>
<tr>
<th>Functioning Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>DENVER II Total</td>
</tr>
</tbody>
</table>

Additional Comments:
References


American Orthopsychiatric Association, 55(4), 603-609.