AN ABSTRACT OF THE THESIS OF

Gregory L. Page for the Master of Science

in Psychology presented on April 9, 1991

Title: Relationships Between Scores on the Wechsler Preschool and Primary Scale of Intelligence-Revised and the Peabody Picture Vocabulary Test-Revised

Abstract Approved: Cooper B. Holmes

The importance of individual intellectual assessment for academic placement requires valid instrumentation. One method of ensuring a valid instrument is to revise an intelligence test. This step was performed to produce the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R) in 1989.

The present study was designed to establish concurrent criterion-related validity of the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R) by comparing it with the Peabody Picture Vocabulary Test-Revised (PPVT-R). Fifty-one children (28 boys and 23 girls) ranging in age from 3 years to 7 years were tested with both instruments.

The WPPSI-R yielded four scores (Vocabulary Subtest, Verbal IQ, Performance IQ, and Full Scale IQ) while the PPVT-R produced one (Standard Score Equivalent). The mean scores obtained in this study were above the normative samples for both tests. Correlation coefficients between the PPVT-R and WPPSI-R scores were low. WPPSI-R Verbal IQ

the PPVT-R and WPPSI-R scores were low. WPPSI-R Verbal IQ and PPVT-R SSE ($r$ = .36) and WPPSI-R Full Scale IQ and PPVT-R SSE ($r$ = .34) were the only significant correlations ($p$<.01). A difference exists for males between the WPPSI-R Performance IQ and PPVT-R Standard Score Equivalent. Some extraneous variables are discussed as possible explanations for the diminished correlations. Further research is needed to confirm or refute these conclusions.

# Relationships Between Scores on the Wechsler Preschool and Primary Scale of Intelligence-Revised and the Peabody Picture Vocabulary Test-Revised

---

A Thesis

Presented to

The Division of Psychology and Special Education

Emporia State University

---

In Partial Fulfillment

of the Requirements for the Degree

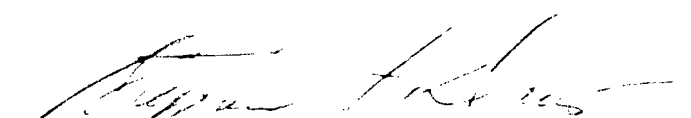
Master of Science

---

by

Gregory L. Page

May, 1991



Approved for the Major Division

James L Wolf

Approved for the Graduate Council

ACKNOWLEDGEMENTS

I would like to sincerely thank Dr. Cooper B. Holmes for his support and guidance throughout this project. I would also like to extend immense gratitude to Mr. Howard H. Carvajal for his benevolent support and assistance during this study. I would also like to express thanks to Dr. Phillip Wurtz for his assistance as a thesis committee member.

I would also like to thank Bob Logan for his assistance in administering tests. An additional thanks goes to Dr. Jim Parks for his two-fold assistance, assisting in test administration and statistical analysis. A heart-felt thank you goes to my parents, Wm. G. and Norma Page for their love, support, guidance, and assistance in my educational conquests. A final thank you goes to Larry Roth for his motivational help in completing this project.

TABLE OF CONTENTS

LIST OF TABLES

# Chapter 1

## Introduction

Individual assessment has been an integral part of psychology from its beginning. Generally, the early formal assessment of humans focused on behavior and sensory reactions, as demonstrated by Wundt and Watson. These experiments took place in as highly controlled environments as possible, such as a specifically designed laboratory with finely calibrated instruments to ensure accurate results. Out of this experimental method, the psychometric movement developed.

Tests that have developed from this psychometric movement include intelligence tests. Some commonly used tests in this category include the Wechsler Intelligence Scales, the Stanford-Binet Test, and the Peabody Picture Vocabulary Test-Revised. These tests of intelligence and ability are held in high regard within the field of psychology, especially in clinical and educational settings (Lubin, Larsen, & Matarazzo, 1984).

A look at definitions of intelligence is necessary to understand intelligence testing fully, as there are numerous definitions of intelligence (Anastasi, 1988). The word intelligence is plagued by a diversity of meanings, not only among the lay public but within experts in a variety of fields, such as psychology, biology, philosophy, and education. Definitions of intelligence are based in part

from memory. This approach views intelligence as stemming from a combination of a genetic component as well as environmental situations (Wechsler, 1989).

The final approach, psychometric theory, emphasizes individual differences of intelligence. The field of intelligence testing finds its roots in this approach. This approach has many theorists that stand out in the field of psychological testing such as Binet, Spearman, Thorndike, Thurstone, Cattell, and Wechsler. Each theorist was involved in the complex history of the development of intelligence testing. The main similarity between the theorists was they believed intelligence could be assessed with a test that measured a factor or combination of factors that defined intelligence. Another prominent theme among these theorists was intelligence as a global performance of the human being. The intelligence tests that developed from this approach report scores as intelligence quotients (IQs) which are nothing more than an expression of the individual's ability at a given time, relative to age norms presented within the test manual (Anastasi, 1988).

The fact that intelligence is a multi-defined term allows for multiple testing measures to be produced. Each theorist accepts the test developed by a previous theorist to be adequate for assessment or creates a new test to probe intelligence as his or her theory postulates. Research, some of which will be discussed below, has shown

correlations between many of the same qualities of intelligence assessed by different methods. This type of research, known as concurrent criterion-related validity, is considered to be an acceptable manner to evaluate a new test or revision of a test.

## Review of the Literature

The literature reviewed below examines a variety of information. It focuses upon intelligence theories and development as well as two published instruments which measure intellectual capacity.

### Theories and Development of Intelligence

Prior to Spearman and others mentioned below, Alfred Binet developed a measure of intelligence that would be the forerunner of future intelligence tests. Binet, with Theodore Simon, constructed the Binet-Simon Scale (1905) to assess an approximate level of each child's intellectual development. Binet and Simon based their scale on the belief that intelligence was a single characteristic that could be measured by questions that required judgement, reasoning, and problem-solving.

The psychometric movement developed early in the twentieth century with the help of the prominent theorist, Charles Spearman. Spearman proposed a theory that centered on a general factor (g) of intelligence. He originally proposed that intelligence could be reduced to one common factor. He later proposed a two-factor theory of

intelligence which was composed of a g factor, a single causal factor of intelligence, and a specific factor (s), a specific ability measured by a test. In Spearman's theory many s's were possible along with g. The g factor could not be reduced to a composite of s's, which resulted in lowered correlations between intelligence tests (Matarazzo, 1972). Another theorist and developer of one of the most popular scales of intelligence, David Wechsler, referred to himself as "a reformed but unchastened Spearmanite" (Wechsler, 1944, p. 3).

Staats and Burns (1984) introduced a theoretical approach of intellectual development called social-behaviorism theory. They viewed intelligence as a process that evolved through an individual's experience. The individual must acquire one necessary skill of an intellectual task before the next one can be gained. In summary, intelligence consists of specific repertoires or systematic skills which are learned in an orderly manner. These repertoires entail language-cognitive skills, sensorimotor skills, and emotional-motivational skills, considered to be basic behavioral repertoires. The acquisition of a later skill is contingent upon achieving an earlier skill. The repertoires of intelligence are called basic behavioral repertoires, in that they determined how well the individual learns and solves problems and how intelligent the individual is in a diversity of settings.

This theory views intelligence as a product of environmental forces because intelligence could be improved with proper stimulation and environment. The authors of this theory reported the cause of intelligence had not been researched previously.

Research discussed in several articles as well as psychological testing textbooks has stated characteristics of intellectual development. A standard growth curve of intellectual development was formulated. It showed that intelligence grows rapidly until the age of twelve when it slows slightly until it reaches asymptote at the age of twenty-two when it begins to slowly decrease with age (Carrol, Kohlberg, & DeVries, 1984).

Wilson (1975) studied cognitive development of 142 pairs of twins aged 4 through 6 years with the Wechsler Preschool and Primary Scale of Intelligence (WPPSI). The results indicated that by the age of six, the child's cognitive skills were functionally related to those of adult intelligence. It was concluded that genetics played a significant role in intellectual development. The IQs of monozygotic twins correlated significantly higher than dizotic twins across all scores reported.

A pair of Soviet researchers (Karpov & Talyzina, 1986) observed intellectual development of preschool age children. The measurement used was an unspecified task that utilized cards with figures drawn on them and shapes similar to the

figures. The children were required to match the shapes with the figures on the cards. The results showed that the children differed on the first trials of each new task and progressed at different rates of carrying out tasks.

Humphreys and Davey (1988) reanalyzed data from Wilson's Louisville Twin Project (1983) which studied the intellectual growth of children from 12 months to 9 years of age. The data were collected through the use of the Bayley, Stanford-Binet, McCarthy Scales of Children's Abilities, WPPSI, and the Wechsler Intelligence Scale for Children (WISC). The findings showed that intelligence was unstable until the age of two years and that a predictive value of IQ did not emerge until late preschool age. A review of older studies prepared by Dunst and Rheingrover (1981) supported the findings reported by Humphreys and Davey.

Theoretical work has produced numerous instruments to assess intelligence such as the Wechsler Primary and Preschool Scale of Intelligence-Revised (WPPSI-R) (1989) and the Peabody Picture Vocabulary Test-Revised (PPVT-R) (1981). Since the present study utilizes the WPPSI-R and the PPVT-R, they will be reviewed in detail below.

Development of the Peabody Picture Vocabulary Test-Revised

The Peabody Picture Vocabulary Test-Revised (PPVT-R) (Dunn & Dunn, 1981), which tied for seventh place in frequency of mention of use in a survey of psychiatric hospitals, community mental health centers, schools for the

retarded, counseling centers, and Veterans Administration hospitals (Lubin et al., 1984), was developed to measure a subject's receptive vocabulary for Standard American English. The PPVT-R includes two forms, Form L and Form M as compared to Form A and Form B from the original PPVT. The test was revised from the original PPVT, created in 1951, to gain renormative statistics and to update the test materials. From the original PPVT, many features were retained. Briefly, the retained features were: (1) individual administration; (2) two forms; (3) bold line drawings; (4) wide application; (5) quick administration; (6) untimed; (7) no reading (8) permissible gestured responses; (9) objectivity; (10) norm-referenced interpretation; (11) variable starting points; and (12) no extensive training to administer. A slight change from the original test was that the PPVT-R discontinued the reporting of Connotative IQ and Mental Age Scores. The scores were replaced with Standard Score Equivalent (SSE) and Age Equivalent scores, respectively. The PPVT-R has a mean of 100 with a standard deviation of 15 (Dunn & Dunn, 1981).

The PPVT-R was standardized with a stratified sample of 4,200 children aged 2 years 6 months to 18 years 11 months, as well as 828 adults aged 19 years to 40 years according to 1970 census data. The criteria used to attain representative data included chronological age and gender balance, geographic region, occupation, and ethnic

representation. The normative sample was divided into 100 females and 100 males for every six month age interval from age 2 years 6 months to 6 years 11 months and 100 females and 100 males for every year interval from age 7 years to 18 years 11 months. This was done so that age reference groups could be reported for interpretive value (Bracken, Prasse, & McCallum, 1984; Dunn & Dunn, 1981; Kipps & Hanson, 1983).

The amount of time required to administer the test varies from 10 to 15 minutes. Materials contained in the test are five training plates and 175 test items. The test items, clear, bold lined drawings, are contained in an easel book to aid in administration. The items were reviewed by many researchers to evaluate the racial, sexual, and ethnic balance. The test items were aligned in an ascending order of difficulty with separate starting points by age provided for motivation of subjects (Dunn & Dunn, 1981).

Reliability and Validity of the PPVT-R

Three reliability coefficients of the 2 year 6 month through 18 year 11 month age groups were reported in the manual of the PPVT-R. Split-half reliabilities ranged from .67 to .88 on Form L and .61 to .86 on Form M. Immediate retest alternate forms reliability coefficients ranged from .73 to .91 for raw scores with a median of .82 and for standard scores ranged from .71 to .89 with a median of .79. Delayed retest alternate-forms reliability coefficients ranged from .52 to .90 with a median of .78 for raw scores

and for standard scores ranged from .54 to .90 with a median of .77 (Dunn & Dunn, 1981).

Gifted sixth grade students (Karnes, McCallum, & Bracken, 1982) and preschool children (Bracken & Prasse, 1983) were administered both Form L and Form M of the PPVT-R. The comparability of the two forms was assessed for each group of subjects. The gifted students' mean scaled scores for Form L were approximately four points higher than Form M while the preschool children scores were less than one point different. Forms L and M correlated at .70 and .84 for the gifted students and preschool children, respectively. The coefficients were high enough that the authors considered the two forms comparable.

The stability of the PPVT-R was examined by researchers after it was published. Naglieri and Pfeiffer (1983) demonstrated the stability of the PPVT-R when they tested 29 mentally retarded students that ranged in age from 5 to 12 with the PPVT-R as a pre- and posttest over a seven month interval. No treatment plan was introduced between tests. Bracken and Murray (1984) tested the stability of the PPVT-R with 29 first through fifth grade students enrolled in regular education classrooms. The tests were administered eleven months apart. Both studies reported correlations above .80 between the different test sessions. The stability of PPVT-R scores was demonstrated successfully by the different studies.

The manual for the PPVT-R reported a study conducted in 1979 to measure similarities between Form A of the PPVT and Form L of the PPVT-R. The tests were administered to 1,849 individuals age 3 years to 18 years. The median SSE on the PPVT-R was 7 to 8 points lower than the median scores of the PPVT IQ (Dunn & Dunn, 1981).

Two independent studies also reported that the PPVT-R produced lower scores than the PPVT. Karnes, McCallum, and Bracken (1982) administered both forms of the PPVT and the PPVT-R to 36 sixth grade children classified as gifted students. They found that the mean PPVT IQ scores were higher than the SSE of the PPVT-R Forms L and M by approximately 8 and 4 points, respectively. Hardman and Smith (1984) replicated the design of the previous study but used 32 individuals, 16 diagnosed as developmentally dyslexic or hyperkinetic and 16 students with class grades of A-B from a highly academic private school. The group of dyslexic/hyperkinetic showed a 15.6 point difference between the PPVT and PPVT-R scores while the undiagnosed students showed a 3 point difference between tests. Despite the lower scores reported by the PPVT-R, the two studies recommended the use of the PPVT-R. The updated materials and norms justified the use of the newer test.

No validity coefficients were reported in the PPVT-R manual (Dunn & Dunn, 1981). The validity of the test was measured by researchers after the PPVT-R was

published. The validity studies used concurrent criterion-related validity which compared the PPVT-R with established tests which measured similar qualities.

Bracken and Prasse (1983) administered the PPVT-R and the McCarthy Scales of Children's Abilities to 35 preschool children identified at risk at birth due to prematurity. The age range of the subjects was from 47 to 58 months. Correlation coefficients reported between PPVT-R Form L and McCarthy ranged from .41 to .63 with coefficients from .42 to .69 for Form M of the PPVT-R. The authors felt the scores could be compared well enough which justified the validity of the PPVT-R.

The Peabody Individual Achievement Test (PIAT) was also used to assess the validity of the PPVT-R. Two studies (Bracken & Murray, 1984; Naglieri & Pfeiffer, 1983) administered the PPVT-R as a pre- and posttest, with the PIAT administered as the posttest. Bracken and Murray (1984) reported correlations ranging from .27 to .75 for the pretest and the PIAT, with higher correlations (.21 to .85) between the posttest and the PIAT. Naglieri and Pfeiffer (1983) showed that the scores reported from the PPVT-R Form L and the PIAT varied less than 4 points for all scores. As mentioned above, both studies reported the PPVT-R was stable. The two studies agreed that the higher correlations of the posttest PPVT-R and the PIAT provided validity evidence for the PPVT-R. Achievement tests empirically

demonstrate that they measure the same features as intelligence tests which allow results of either test to correlate (Anastasi, 1988).

Fifty-one elementary and middle school aged children (Hollinger & Sarvis, 1984) and 101 students in special education programs (Worthing, Phye, & Nunn, 1984) were administered the Wechsler Intelligence Scale for Children-Revised (WISC-R) and the PPVT-R. The PPVT-R SSEs were compared to the three IQs, Performance, Verbal, and Full Scale, of the WISC-R. The scores of the PPVT-R compared better with the Verbal and Full Scale IQs than the Performance IQ. The two studies showed a range of correlations of .69 to .82 with comparisons of the PPVT-R and WISC-R Performance IQ the lowest and the PPVT-R and WISC-R Full Scale IQ being the highest correlations. Concurrent criterion-related validity was demonstrated by both studies.

The PPVT-R Form L was compared to the Stanford-Binet IV (Binet IV), Columbia Mental Maturity Scale (CMMS), and Goodenough-Harris Drawing Test (GHDT) by testing 23 third grade children that ranged in age from 8 years 2 months to 9 years 10 months (Carvajal, McVey, Sellers, Weyand, & McKnab, 1987). The PPVT-R correlated significantly only with the Binet IV Composite Standard Age Score (SAS) and Binet IV Vocabulary SAS with coefficients of .60 and .53, respectively. The PPVT-R did not correlate well with the

CMMS and GHDT, with coefficients of .23 and .25, respectively. The PPVT-R correlated well with the Binet IV, an intelligence scale, justifying its use as a screening instrument although it did not correlate well with other screening instruments.

Carvajal, Hardy, Harmon, Sellers, and Holmes (1987) demonstrated the merit of the PPVT-R Form L as a screening instrument of intelligence when related to the Binet IV and CMMS. Twenty-one kindergarten students were tested. They ranged in age from 5 years 1 month to 6 years 11 months. The PPVT-R correlated well with the Binet IV Composite SSE and Vocabulary SSE with coefficients of .56 and .60 respectively. The PPVT-R correlated with a coefficient of .22 with the CMMS. The results support the Carvajal et al. (1987) findings that the PPVT-R was an acceptable screening instrument for intelligence.

The PPVT-R Form L and Binet IV were correlated using 32 college students with a mean age of 18 years 11 months (Carvajal, Gerber, & Smith, 1987). The study reported that no significant mean differences were present between male and female subjects. The Binet IV Composite SAS and PPVT-R SSE correlated at .69, which demonstrates the validity of the PPVT-R.

Carvajal, Shaffer, and Weaver (1989) demonstrated the usefulness of the PPVT-R Form L with 29 male maximum security inmates who were administered the Wechsler Adult

Development of the Wechsler Preschool and Primary Scale of Intelligence-Revised

The Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R) was developed to update WPPSI (Wechsler, 1967) norms and to extend the age range of the scale from 4 years through 6 years 7 months to three years through 7 years 3 months. This essentially extended the age range 1 year downward and 1 year upward. The WPPSI-R was intended to measure intellectual ability in educational, clinical, and research settings. The primary intent was to identify intelligent children and secondly to assess change in performance of individuals over time which occurred naturally or was manipulated.

The test was administered to 2,100 children age 3 years to 7 years 3 months in a stratified fashion based upon data from the Bureau of the Census in 1986. Factors considered in the stratification were age, sex, geographic region, ethnicity, and parental education and occupation. Of the original 2,100 children, 1,700 were used for normative information, with the remaining 400 used to investigate time bias toward minority subjects. Nine age groups were derived from the age ranges reported above at half year intervals. Within each age range, 100 boys and 100 girls were tested, except the 7 year through 7 year 3 month age group which included 50 boys and 50 girls (Wechsler, 1989).

As with all other Wechsler scales, the WPPSI-R

continues to use a variety of subtests that measure two factors of intelligence: Verbal and Performance. Eleven subtests were included in the original WPPSI. Five subtests comprised Performance IQ: Animal House, Picture Completion, Mazes, Geometric Design, and Block Design. Six subtests comprised Verbal IQ: Information, Vocabulary, Arithmetic, Similarities, Comprehension, and Sentences. The Sentences subtest was a supplementary test and was not used in determination of the IQ. The subtests were downward extensions of subtests on the Wechsler Intelligence Scale for Children (WISC). Subtests on the WPPSI were written at an easier level for younger children. The WPPSI-R retained the eleven subtests, added a new one, Object Assembly, and changed Animal House to Animal Pegs. The use of the supplementary scales, Animal Pegs and Sentences, was also retained from the WPPSI. The subtests are administered in alternating order of type; one Performance, one Verbal, etc. All subtests are administered to each individual. Different starting points are utilized to ensure individuals are not administered a succession of too many easy items. Varied cut off or ending conditions are also used as not to overwhelm the individual with too many difficult items.

Each subtest measures different areas of a person's mental abilities. The order and qualities measured by each subtest outlined below are: Object Assembly, constructive imagination; Information, background information obtained by

the individual from home and school; Geometric Design, visual-motor perceived coordination; Comprehension, common sense, practical judgement, and self-direction; Block Design, versatility of problem solving, concept formation, and spatial abilities; Arithmetic, numerical reasoning and power of concentration; Mazes, planning, foresight, visual-motor coordination, and speed in conjunction with accuracy; Vocabulary, language background, defining words, and verbal ability; Picture Completion, visual alertness, awareness of detail, and power of observation; Similarities, generalized, categorical thinking and abstract verbal reasoning; Animal Pegs, memory, attention span, and associative abilities; and Sentences, span of attention and immediate auditory recall (Wechsler, 1958, 1989).

Each subtest produces a raw score which is converted to a scaled score derived from the norm tables of the standardization group. The subtests report a mean of 10 and a standard deviation of 3. Three separate IQ scores are reported: Performance, Verbal, and Full Scale with a mean of 100 and a standard deviation of 15. The IQs were developed to assess the two factors presented within the test as well as the overall intellect of the individual. Performance and Verbal IQ scores are determined by the summated scaled scores presented by the subtests of the appropriate factor. Full Scale IQ is derived from the summed scaled scores of the Verbal and Performance IQs.

Reliability and Validity of the WPPSI-R

Split-half average estimated reliability coefficients are reported in the WPPSI-R manual (Wechsler, 1989). The values were corrected by the Spearman-Brown formula which resulted in the estimated coefficients. The coefficients for the Performance subtests ranged from .63 (Object Assembly) to .85 (Block Design and Picture Completion). The Verbal subtest coefficients ranged from .80 (Arithmetic) to .86 (Similarities). The coefficients for the three IQs were: Performance IQ, .92; Verbal IQ, .95; and Full Scale IQ, .96.

The stability of the scales was tested on 175 children from the standardization group after a seven week interval. Stability coefficients reported for Performance subtests ranged from .52 (Mazes) to .82 (Picture Completion); from .70 (Similarities) to .81 (Information) for Verbal subtests; .88 for Performance IQ; and .90 for Verbal IQ; and .91 for Full Scale IQ.

A validity study was reported in the WPPSI-R manual comparing the WPPSI-R with the WPPSI with 144 children aged 48 to 79 months. The study was conducted by the publishers of the test. The subjects were not used in the standardization group. The results showed correlations for the Performance, Verbal, and Full Scale IQs at .82, .85, and .87, respectively (Wechsler, 1989).

Other validity studies were reported with the WISC-R,

Stanford-Binet Intelligence Scale: Fourth Edition (Binet IV), McCarthy Scales of Children's Abilities (MSCA), and Kaufman Assessment Battery for Children (K-ABC) (Wechsler, 1989). Fifty children aged 72 to 86 months were administered the WPPSI-R and the WISC-R at a 7 to 38 day interval. The correlation coefficients reported ranged from .55 (WPPSI-R Performance IQ and WISC-R Verbal IQ) to .85 (WPPSI-R and WISC-R Full Scale IQs). The WPPSI-R reported IQ scores 5 to 9 points lower than the WISC-R IQ scores. The validity of the WPPSI-R was supported as an accepted measure for intelligence.

The WPPSI-R and Binet IV were administered to 115 children who ranged in age from 4 to 7 years. The three IQ scores from the WPPSI-R were correlated with the Binet IV Composite Standard Age Score with moderate coefficients that ranged from .56 (Performance) to .74 (Full Scale). The Binet IV composite was approximately 2 IQ points higher than the three WPPSI-R IQs.

The MSCA General Cognitive Index (GCI) was compared to the 3 IQ scores of the WPPSI-R. Ninety-three children aged 4 to 6 were administered the tests. Moderate correlation coefficients were reported which ranged from .66 (Performance) to .81 (Full Scale). The WPPSI-R scores were approximately 2 IQ points lower than the MSCA GCI.

The final study reported in the WPPSI-R manual compared the K-ABC with the WPPSI-R. THe subjects tested were fifty-

nine children aged 37 to 76 months. Low to moderate correlations were reported between the two tests. The highest correlation was .49 between the WPPSI-R Full Scale IQ and the K-ABC Mental Processing Composite. The low correlations were attributed to significant differences of sample composition and/or that the two scales measured different constructs.

The studies presented above, except the WPPSI-R/K-ABC comparison, supported the concurrent criterion-related validity of the WPPSI-R. It was concluded that the WPPSI-R measured intelligence comparable to other tests with slightly lower scores. The results supported the use of the WPPSI-R to measure intelligence of preschool and elementary aged children.

Purpose of the Study

The purpose of this study was to research the validity of the WPPSI-R by examining the relationships between the WPPSI-R and the PPVT-R. The WPPSI-R has been favorably reviewed by the authors of the test, but the results may be biased. The PPVT-R was chosen because it is the most popular screening test available, and it has not been correlated with the WPPSI-R by the publisher. It is in this light that the validity of the WPPSI-R could be measured by unbiased researchers.

Statement of Significance

The results of this study will enable psychologists who

use the WPPSI-R to make an additional judgement of its validity. The PPVT-R has been an acceptable screening instrument for years, and its relationship with the WPPSI-R should be of paramount importance to school psychologists and those professionals who work with children.

# CHAPTER 2

# METHOD

## Subjects

This study included 51 children (28 boys and 23 girls) from a midwestern city with a population of approximately 27,000. The sample included students ranging in age from 3 years 6 months to 7 years 3 months with a mean age of 5 years 7 months (SD = 1 year 0 months). The subjects were classified by yearly age categories. The children were obtained by contacting Kansas Unified School District 253. The school district was asked for permission to seek parental consent for child volunteers.

The parents of the children were asked to read and sign an informed consent form before data were collected. The form described the intent and purpose of the study, the procedures of the study, and the rights of the child. The signed form verified an agreement for the child to be tested. Confidentiality was ensured through standard techniques. Approval to conduct the study was obtained from to Emporia State University's Review Board for Treatment of Human Subjects.

## Procedure

The WPPSI-R and the PPVT-R were individually administered to each subject. The two tests were administered in a counterbalanced sequence with the data being collected over a three month period. The age, sex,

and name of the subject were collected to allow proper scoring.

The WPPSI-R was individually administered by either of two males in rooms specifically designed for testing at Emporia State University. The test was administered and scored according to the procedures outlined in the manual.

The PPVT-R was individually administered by a third male in the same fashion as the WPPSI-R, in strict accordance with the manual. The raw scores were converted to standard scores with the use of tables included in the test manual.

Statistical Design

The type of research described in this study was correlational research. The study examined how two variables, the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R) (Wechsler, 1989) and Peabody Picture Vocabulary Test-Revised (PPVT-R) (Dunn & Dunn, 1981), related without manipulating either variable. This was done by computing correlation coefficients between the scores each test yielded.

The research described above is a concurrent criterion-related validity study of the WPPSI-R. Concurrent criterion-related validity research is done by comparing a new test or instrument with an older, established instrument. This study compared the WPPSI-R with the PPVT-R.

The study produced five scores per subject: four scores from the WPPSI-R (Vocabulary subtest, Verbal IQ, Performance IQ, and Full Scale IQ) and one from the PPVT-R (Standard Score Equivalent). These scores were analyzed by calculating the Pearson product-moment correlation coefficient (Pearson *r*) to determine the association between the WPPSI-R and PPVT-R scores. The Pearson *r* has been extensively used in previous studies of concurrent criterion-related validity of intelligence tests. A 2 x 4 factorial analysis of variance (ANOVA) was calculated to determine gender differences of test scores. The mean, mode, and median, range, minimum scores, maximum scores, and standard deviation, for the entire sample and gender groups were also analyzed by comparing them with the normative groups.

# CHAPTER 3

## RESULTS

Four scores from the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R) and one score from the Peabody Picture Vocabulary Test-Revised were obtained for 51 children (28 boys and 23 girls). The means, standard deviations, ranges, minimum scores, maximum scores, modes, and medians are presented in Table 1.

Table 1

Descriptive Statistics for the WPPSI-R and PPVT-R.

| Test | M | SD | Range | Mode | Mdn |
|---|---|---|---|---|---|
| WPPSI-R | | | | | |
| Vocab | 11.43 | 3.17 | 12 (4-16) | 13 | 12 |
| Ver IQ | 107.64 | 14.45 | 76 (69-145) | 97 | 107 |
| Per IQ | 108.82 | 13.21 | 65 (80-145) | 97 | 109 |
| FS IQ | 109.49 | 14.05 | 68 (74-142) | 101 | 108 |
| PPVT-R | | | | | |
| SSE | 110.98 | 12.24 | 57 (79-136) | 105 | 110 |

The mean IQ scores reported by this study range from 107.64, SD = 14.45 (Verbal) to 109.49, SD = 14.05 (Full Scale). These mean IQ scores are somewhat higher than the normative groups for the WPPSI-R which report a mean of 100, SD = 15. The Vocabulary subtest from the WPPSI-R also

yields a higher mean score (11.43, SD = 3.17) than the normative sample (10, SD = 3). The mean PPVT-R SSE score (110.98, SD = 12.24) is also higher than the normative group which reports the same statistic as the WPPSI-R. Tables 2 and 3 list the descriptive statistics of the subjects by gender.

Table 2

Descriptive Statistics for the WPPSI-R and PPVT-R for Males.

| Test | M | SD | Range | Mode | Mdn |
|---|---|---|---|---|---|
| WPPSI-R | | | | | |
| Vocab | 11.86 | 3.49 | 12 (4-16) | 15 | 12.5 |
| Ver IQ | 110.43 | 16.91 | 76 (69-145) | 97 | 109.5 |
| Per IQ | 106.75 | 11.50 | 49 (80-129) | 106 | 107.5 |
| FS IQ | 110.11 | 14.75 | 62 (74-136) | 101 | 107.5 |
| PPVT-R | | | | | |
| SSE | 113.36 | 12.15 | 48 (88-136) | 105 | 116.0 |

Table 3

Descriptive Statistics for the WPPSI-R and PPVT-R for Females.

| Test | M | SD | Range | Mode | Mdn |
|---|---|---|---|---|---|
| WPPSI-R | | | | | |
| Vocab | 10.91 | 2.71 | 92 (7-16) | 9 | 10.0 |
| Ver IQ | 104.26 | 10.08 | 42 (84-126) | 107 | 103.0 |
| Per IQ | 111.35 | 14.93 | 55 (90-145) | 97 | 109.0 |
| FS IQ | 108.74 | 13.43 | 56 (86-142) | 94 | 108.0 |
| PPVT-R | | | | | |
| SSE | 108.09 | 11.97 | 54 (79-133) | 105 | 108.0 |

The mean IQ scores reported in Table 2 are above the normative samples for the WPPSI-R. Males performed better than females in regards to mean scores on the WPPSI-R Vocabulary Subtest, Verbal IQ, Full Scale IQ, and PPVT-R SSE, while the females scored higher on the mean Performance IQ. The male mean IQ scores range from 106.75, SD = 11.50 (Performance) to 110.43, SD = 16.91 (Verbal). The female mean IQ scores range from 104.26, SD = 10.08 (Verbal) to 111.35, SD = 14.93 (Performance).

Pearson product-moment correlation coefficients were calculated between the WPPSI-R and PPVT-R. One-tailed significance was also calculated for the scores. Table 4

# REFERENCES

Anastasi, A. (1988). Psychological Testing (6th ed.). New York: Macmillan.

Bracken, B. A., & Murray, A. M. (1984). Stability and predictive validity of the PPVT-R over an eleven month interval. Educational and Psychological Research, 4, 41-44.

Bracken, B. A., & Prasse, D. P. (1983). Concurrent validity of the PPVT-R for "at risk" preschool children. Psychology in the Schools, 20, 13-15.

Bracken, B. A., Prasse, D. P., & McCallum, R. S. (1984). Peabody Picture Vocabulary Test-Revised: An appraisal and review. School Psychology Review, 13, 49-60.

Carrol, J. B., Kohlberg, L., & DeVries, R. (1984). Psychometric and Piagetian intelligences: Toward resolution of controversy. Intelligence, 8, 67-91.

Carvajal, H., Gerber, J., & Smith, P. D. (1987). Relationship between scores of young adults on Stanford-Binet IV and Peabody Picture Vocabulary Test-Revised. Perceptual and Motor Skills, 65, 721-722.

Carvajal, H., Hardy, K., Harmon, K., Sellers, T. A., & Holmes, C.B. (1987). Relationships among scores on the Stanford-Binet IV, Peabody Picture Vocabulary Test-Revised, and Columbia Mental Maturity Scale. Bulletin of the Psychonomic Society, 25, 275-276.

Carvajal, H., McVey, S., Sellers, T., Weyand, K., & McKnab, P. (1987). Relationships between scores on the general purpose abbreviated battery of Stanford-Binet IV, Peabody Picture Vocabulary Test-Revised, Columbia Mental Maturity Scale, and Goodenough-Harris Drawing Test. The Psychological Record, 37, 127-130.

Carvajal, H., Shaffer, C., & Weaver, K. A. (1989). Correlations of scores of maximum security inmates on Wechsler Adult Intelligence Scale-Revised and Peabody Picture Vocabulary Test-Revised. Psychological Reports, 65, 268-270.

Dunn, L. M., & Dunn, L. M. (1981). Manual for the Peabody Picture Vocabulary Test-Revised. Circle Pines, MN: American Guidance Service.

Dunst, D. J., & Reingrover, R. M. (1981). Discontinuity and instability in early development: Implications for assessment. Topics in Early Childhood Special Education, 1, 49-60.

Hardman, P. K., & Smith, D. D. (1984). The PPVT-R: Is it perceptually biased? Academic Therapy, 19, 373-377.

Hollinger, C. L., & Sarvis, P. A. (1984). Interpretation of the PPVT-R: A pure measure of verbal comprehension? Psychology in the Schools, 21, 34-41.

Humphreys, L. G., & Davey, T. C. (1988). Continuity in intellectual growth from 12 months to 9 years. Intelligence, 12, 183-197.

Karnes, F. A., McCallum, R. S., & Bracken, B. A. (1982). Comparison of the PPVT and PPVT-R as possible instruments for screening gifted children. Psychological Reports, 51, 591-594.

Karpov, Y. V., & Talyzina, N. F. (1986). Criteria of intellectual development in children. Soviet Psychology, 24, 77-88.

Kipps, D., & Hanson, D. (1983). Test review: The revised PPVT. School Psychology Review, 12, 112-113.

Kutsick, K., Vance, B., Schwarting, F. G., & West, R. (1988). A comparison of three different measures of intelligence with children identified at-risk. Psychology in the Schools, 25, 270-275.

Lubin, B., Larsen, R. M., & Matarazzo, J. D. (1984). Patterns of psychological test usage in the United States: 1935-1982. American Psychologist, 39, 451-454.

Matarazzo, J. D. (1972). Wechsler's Measurement and Appraisal of Adult Intelligence (5th ed.). New York: Oxford University Press.

Naglieri, J. A., & Pfeiffer, S. I. (1983). Stability, concurrent and predictive validity of the PPVT-R. Journal of Clinical Psychology, 39, 965-967.

Staats, A. W., & Burns, G. L. (1984). Intelligence and child development: What intelligence is and how it is learned and functions. Genetic Psychology Monographs, 104, 237-301.

Wechsler, D. (1944). The Measure of Adult Intelligence (3rd ed.). Baltimore: The Williams & Wilkins Company.

Wechsler, D. (1958). The Measurement and Appraisal of Adult Intelligence (4th ed.). Baltimore: The Williams & Wilkins Company.

Wechsler, D. (1967). Manual for the Wechsler Preschool and Primary Scale of Intelligence. New York: The Psychological Corporation.

Wechsler, D. (1989). Manual for the Wechsler Preschool and Primary Scale of Intelligence-Revised. San Antonio: The Psychological Corporation.

Wilson, R. S. (1975). Twins: Patterns of development as measured on the Wechsler Preschool and Primary Scale of Intelligence. Developmental Psychology, 11, 126-134.

Wilson, R. S. (1983). The Louisville twin study: Development synchronies in behavior. Child Development, 54, 198-216.

Worthing, R. J., Phye, G. D., & Nunn, G. D. (1984). Equivalence and concurrent validity of PPVT-R Forms L and M for school-age children with special needs. Psychology in the Schools, 21, 296-299.

: All Graduate Students Who Submit a Thesis or Research Problem/Project as Partial Fulfillment of the Requirements for an Advanced Degree

OM: Emporia State University Graduate School

Gregory L. Page, hereby submit this esis/report to Emporia State University as partial fulfillment the requirements for an advanced degree. I agree that the brary of the University may make it available for use in cordance with its regulations governing materials of this type. further agree that quoting, photocopying, or other reproduction this document is allowed for private study, scholarship ncluding teaching) and research purposes of a nonprofit nature. copying which involves potential financial gain will be lowed without written permission of the author.

Signature of Author

April 23, 1991
Date

Relationship Between Scores on the Wechsler Preschool and Primary Scale of Intelligence-Revised and the Peabody Picture Vocabulary Test-Revised

Title of Thesis/Research Project

Signature of Graduate Office Staff Member

Date Received

stribution: Director, William Allen White Library
Graduate School Office
Author