AN ABSTRACT OF THE THESIS OF

Gina L. Lomax for the Master of Science

in Psychology presented on April 15, 1992

Title: Relationships Between Scores on the Peabody Picture Vocabulary Test-Revised and the McCarthy Scales of Children's Abilities

Abstract Approved: Cooper B. Holmes

Diagnosticians involved in the assessment of children's intelligence need to be provided with information on the validity of testing instruments as well as the specific abilities that are assessed by various tests. Prior to a test gaining acceptance as an appropriate instrument for measuring intelligence, its concurrent validity must be substantiated.

The present study was designed to establish the concurrent criterion-related validity of the Peabody Picture Vocabulary Test-Revised (PPVT-R) by comparing it with the McCarthy Scales of Children's Abilities (MSCA). Thirty-two children (17 boys and 15 girls) ranging in age from 4 years, 2 months to 6 years, 3 months were administered both instruments in the study.

The MSCA yielded four scores (Verbal, Perceptual-Performance, Quantitative and General Cognitive Index) and the PPVT-R produced one score (Standard Score Equivalent). All of the mean scores obtained in the study were somewhat higher than the normative samples

for both instruments. Although the overall mean scores on the two instruments were very similar, the tests correlated at a low level.

A 2 X 2 X 5 mixed factor analysis of variance was conducted to examine differences in gender (Boys, Girls), grade-level (Preschool, Kindergarten) and test scores (Verbal, Perceptual-Performance, Quantitative, General Cognitive Index and PPVT-R Standard Score Equivalents). No significant main effects were established, nor were any of the interactions significant. The limited sample size included in this study may have influenced the outcome of the study. Therefore, further research is needed in this area in the future with a larger, more diverse sample.

Relationships Between Scores on the Peabody Picture Vocabulary Test-Revised and the McCarthy Scales of Children's Abilities

---

A Thesis

Presented to

The Division of Psychology and Special Education

Emporia State University

In Partial Fulfillment

of the Requirements for the Degree

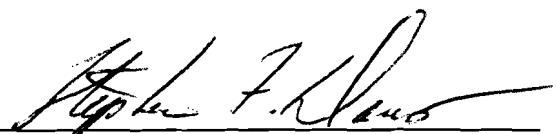
Master of Science

---

by

Gina L. Lomax

May, 1992



Approved for the Major Division

Approved for the Graduate Council

ACKNOWLEDGEMENTS

I would like to sincerely thank Dr. Cooper B. Holmes and Mr. Howard H. Carvajal for all of their support and guidance throughout this project and my graduate program. I feel very honored to have had the opportunity to learn from both of these individuals. I would also like to thank Dr. Philip Wurtz for serving as a thesis committee member.

I would like to thank Dr. Kenneth A. Weaver for his assistance in the statistical analysis of this study. I would also like to thank Becky Hendrix for all of her support and assistance on this project. Finally, a thank you goes to some very special individuals, my family, Deanna Dinges, Chris Smith, Susan Foster and Lynne Daise for all of their support throughout my educational endeavors.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Introduction

Classifying individuals with respect to cognitive abilities has long been a concern of psychologists and educational diagnosticians (Davis & Kramer, 1985). With the increasing recognition of the early years as critical in a child's development, psychologists are evaluating children at younger ages. Belief in the importance of early childhood experience has led to an increasing emphasis on educational programming for both normal and exceptional children (Vance, 1982). Additionally, Public Law 94-142 (Education for All Handicapped Children Act, 1979) requires that states desiring financial assistance from the government implement policies that assure an appropriate and effective public education for all children, regardless of handicap (Harrison & Naglieri, 1981). This law also mandates psychological reevaluations at three-year intervals for children who participate in special education programs (Levenson & Lasher-Adelman, 1988). In addition, an extension to Public Law 94-142 was passed in 1986 entitled Public Law 99-457. Public Law 99-457 (Education of the Handicapped Act Amendment of 1986) established two new federal programs to address special education services for handicapped and "at

risk" children between the ages of birth and six.

Psychologists are continually searching for a time and cost efficient instrument for reliably assessing children's abilities (Taylor, 1979). However, one of the problems facing psychologists is the lack of adequately normed assessment instruments for preschool and kindergarten aged children (Vance, 1982). There is an extensive amount of tests available to assess children's abilities; unfortunately, most such tests are inadequately developed. Also, test authors' descriptions of the validity of the tests have seldom been investigated (Sommers, Erdige & Peterson, 1978).

The search for an accurate screening device has led to a number of studies in which investigators have compared multi-skill tests such as the Wechsler Intelligence Scale for Children (WISC), the Stanford-Binet and the McCarthy Scales of Children's Abilities (MSCA) with single-skill tests such as the Peabody Picture Vocabulary Test (PPVT). Investigators have attempted to determine similarities and differences of the tests as well as the advantages and disadvantages of using multi-skill and single-skill tests interchangeably (Taylor, 1979). The multi-skill tests often require a lengthier administration and interpretation period as compared to the single-skill

tests which require a limited amount of clinical time. Therefore, the single-skill tests are often the preferred test. However, numerous difficulties can be encountered when implementing the results of single-skill tests for clinical, research or educational purposes if they fail to adequately represent an individual's abilities (Ferrari, 1980). It has been suggested that occasions exist in which a rough appraisal of an individual's ability may be required and/or when time available for testing is limited that a single-skill test could accurately supply the needed information (Carvajal, Shaffer & Weaver, 1989). It would appear to be of significant clinical importance to determine the comparability of multi-skill and single-skill tests. Numerous instruments have been developed in order to assess intelligence such as the McCarthy Scales of Children's Abilities (MSCA) (McCarthy, 1972) and the Peabody Picture Vocabulary Test-Revised (PPVT-R) (Dunn & Dunn, 1981). Since the following study focuses on the PPVT-R and the MSCA, 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) was designed as a measure of a

person's receptive language for Standard American English. This test replaced the original Peabody Picture Vocabulary Test (PPVT) published in 1959, retaining many of the features of its predecessor. The PPVT-R consists of two forms: Form L and Form M as compared to Form A and Form B of the 1959 edition of the PPVT. The test is individually administered and typically requires approximately 15 to 20 minutes (Dunn & Dunn, 1981).

Advances in the construction of the revised version of the PPVT allow it to be a more sophisticated instrument than its predecessor. The PPVT-R was standardized on a large representative sample of children and adults from age 2 years, 6 months through adulthood. In assembling the pool of stimulus words, the following 19 categories were established to ensure a proper balance of nouns, verbs and adjectives: (1) actions; (2) animals; (3) buildings; (4) clothing; (5) descriptors; (6) foods; (7) household and yard fixtures; (8) household utensils; (9) human body parts; (10) human workers; (11) human and humanoid forms, such as statues; (12) mathematical terms; (13) plants; (14) produce; (15) school and office supplies; (16) tools and recreational items; (18) vehicles; and (19) weather and geographical items (Dunn & Dunn, 1981).

The revised version of the PPVT was also improved through the virtual elimination of gender and ethnic stereotyping. All of the drawings were reworked to allow for better racial, gender and ethnic balance. Additionally, 25 items were added to each form in order to increase the test's sensitivity. The test items were clear, bold line drawings presented in an easel book for ease of administration (Kipps & Hanson, 1983). The highly criticized terms "IQ" score and "Mental Age" score were replaced with Standard Score Equivalent and Age Equivalent Scores. The PPVT-R has a mean of 100 and a standard deviation of 15 (Dunn & Dunn, 1981).

The manual reports information on the reliability of the test. The split-half reliabilities for all ages ranged from .67 to .88 on Form L and .61 and .86 on Form M. The alternate-forms reliability coefficients based on immediate retest ranged from .73 to .91 for the various age groups. The alternate-forms reliability coefficients based on delayed retest ranged from .52 to .90. The PPVT-R manual provides a limited amount of information dealing with the validity of the instrument; however, the authors reported that the PPVT-R correlates moderately with other tests of scholastic aptitude (Dunn & Dunn, 1981).

Review of the Literature on the Peabody Picture Vocabulary Test-Revised

Several publishers and test constructors have investigated the psychometric properties of the PPVT-R and its comparability to other tests (Kipps & Hanson, 1983). The PPVT-R has been compared to the Wechsler scales in various studies. In 1981, Prasse and Bracken examined the relationship between the PPVT-R and the Wechsler Intelligence Scale for Children-Revised (WISC-R) for 67 educable mentally retarded students. Significant differences were found between the PPVT-R mean standard score and Verbal IQ, Performance IQ, and Full Scale IQ of the WISC-R. The PPVT-R failed to correlate significantly with the WISC-R subtest scales, suggesting the two tests are measuring different abilities.

A comparison was also done between the PPVT-R and the WISC-R with 32 elementary school-aged children referred for learning disabilities. Significant differences were noted between the mean PPVT-R standard score and each of the three Wechsler Intelligence Quotients. Significant correlations were obtained between the PPVT-R and Verbal, Performance, and Full Scale IQs (Breen, 1981).

Wright (1983) administered the PPVT-R and the

WISC-R to a sample of 35 gifted students who were predominantly white and of a middle-class background. The correlation between the PPVT-R and the WISC-R Full Scale IQ was .27. The PPVT-R displayed modest effectiveness in predicting WISC-R IQ scores; therefore, the PPVT-R appears to offer limited effectiveness as a screening measure for this restricted population.

An investigation of the comparability of the PPVT-R and WISC-R for children and youths referred for psychological services was performed in 1983 (Vance, Kitson & Singer, 1983). The subjects were 37 children with a mean age of 10 years, 1 month. The PPVT-R failed to correlate significantly with the WISC-R IQ scales. In addition, the study established that the PPVT-R underestimated the WISC-R Full Scale IQ by approximately 4 to 6 points.

The comparability of the PPVT-R to the WISC-R was also investigated by Breen and Siewart (1983). The PPVT-R and WISC-R were administered to 30 learning disabled and 29 referred students. Form M of the PPVT-R correlated significantly with all WISC-R measures for both groups. The PPVT-R yielded significantly lower scores than the WISC-R for both groups.

In 1984, Rosso, Falasco and Koller reviewed the

relationship between the PPVT, the PPVT-R and the WISC-R with incarcerated delinquents. The subjects were 36 delinquents between the ages of 13 years, 10 months and 16 years, 10 months. The PPVT-R correlated significantly with the WISC-R Verbal IQ, Performance IQ, Full Scale IQ, and PPVT.

An additional study focused on the comparability of the WISC-R subtest scores and the PPVT-R standard score in order to determine abilities measured by the PPVT-R (Hollinger & Sarvis, 1984). This study was conducted on a sample of 51 rural children. Examination of the raw data revealed that the PPVT-R overestimated the WISC-R Full Scale IQ for 21 of the subjects. The subjects differed significantly only in terms of performance on the Arithmetic subtest, suggesting that abilities assessed by the Arithmetic subtest may determine whether the PPVT-R underestimates or overestimates a child's WISC-R Full Scale IQ. This study also suggested that verbal comprehension abilities may contribute most to successful performance on the PPVT-R.

Hollinger and Sarvis (1984) conducted a second study investigating the relationship between the PPVT-R and the WISC-R with 53 rural children. Unlike its predecessor, the PPVT-R underestimated the WISC-R scale

scores, particularly the WISC-R Performance Scale. The PPVT-R was highly correlated with the WISC-R Verbal Scale, Performance Scale, and Full Scale IQ.

Davis and Kramer (1985) evaluated the effect of prior administration of either the PPVT-R or the WISC-R on the scores of the latter administered test. The sample consisted of 40 public school second-grade rural students. Scores from the two tests were moderately correlated. Prior administration of one of the instruments did not appear to alter scores on the PPVT-R or the WISC-R. It was found that normal school children tend to obtain lower scores on the PPVT-R than on the WISC-R.

A study focusing on the relationship of the scores on the PPVT-R and WISC-R with special education children and youths was conducted in 1986 by Candler, Maddux and Johnson. The comparisons were made with 104 children diagnosed as mentally retarded or learning disabled. The PPVT-R underestimated all of the WISC-R Scale IQs.

In 1986, Beck and Black investigated the PPVT-R and the WISC-R using only Form L of the PPVT-R for comparison purposes. A sample of 32 mildly/moderately handicapped students were studied. Significant correlations were obtained between the PPVT-R and Full

Scale IQs, but not the Performance IQ. The PPVT-R mean score underestimated the WISC-R Full Scale IQ by 7 points, supporting the findings of Vance, Kitson and Singer, 1983; Breen and Siewart, 1983; and Hollinger and Sarvis, 1984.

The PPVT-R and WISC-R scores were evaluated in a study of the PPVT-R as a measure of psycholinguistic functioning (Altepeter, 1989). This study involved 75 school-aged children. The mean PPVT-R standard score was not significantly different than the WISC-R Full Scale IQ. In addition, the PPVT-R correlated positively and significantly with the WISC-R Full Scale IQ.

The PPVT-R and WISC-R were examined to determine the presence of sex or racial basis in two studies. First, the role of sex differences incorporated in the PPVT-R and the WISC-R was examined by Smith, Edmonds and Smith (1989). Males consistently scored higher than females on 87% of the dependent variables being measured with the exception of the WISC-R Coding subtest. The PPVT-R was investigated to determine possible racial bias in predicting the scores of white and black students on the WISC-R for 75 adolescent subjects (Halpin, Simpson & Martin, 1990). The PPVT-R was found to be the valid for blacks and whites.

The relationship between scores on the PPVT-R and the Wechsler Adult Intelligence Scale-Revised (WAIS-R) was compared in the following studies. One study focused on the validity of the PPVT-R with mentally retarded adults. The results indicated that the PPVT-R tended to yield significantly lower estimates of intellectual functioning than did the PPVT or WAIS-R. A mean difference beyond 17 points was obtained, which implies the revised version of the PPVT may be inappropriate for use with adult mentally retarded individuals (Prout & Schwartz, 1984).

The second study reviewed correlations of the scores of maximum security inmates on the PPVT-R and the WAIS-R. The subjects were 29 male inmates. A significant correlation was established between the PPVT-R and WAIS-R Full Scale IQ. The authors concluded that the PPVT-R could serve as an effective screening test for this population; however, one should not utilize the tests interchangeably (Carvajal, Shaffer & Weaver, 1989).

Kutsick, Vance, Schwarting and West (1988) compared three different measures of intelligence with 70 preschool children identified as "at risk." The preschoolers were administered the WPPSI, the PPVT-R and the Expressive One-Word Picture Vocabulary Test

(EOWPVT). Results revealed that a significant difference existed between the PPVT-R and the WPPSI Full Scale IQ, with the largest difference of 8 points occurring between the PPVT-R standard score and the WPPSI Performance IQ. The subjects' scores on the PPVT-R and EOWPVT were generally comparable to the WPPSI Verbal IQ score. Similar results were found by Vance, West and Kutsick (1989).

Carvajal, Hardy, Harmon, Sellers and Holmes (1987) examined the relationships between the Stanford-Binet IV, PPVT-R and Columbia Mental Maturity Scale. The sample included 21 kindergarten-aged children. A correlation of .56 was obtained between the Binet IV and PPVT-R. The results of this study were consistent with the median value of .62 derived from 72 studies of the 1972 Stanford-Binet and PPVT (Dunn & Dunn, 1981).

In 1987, Carvajal, McVey, Sellers, Weyand and McKnab investigated the relationships between scores on the general purpose abbreviated battery of Stanford-Binet IV, the PPVT-R, the Columbia Mental Maturity Scale and the Goodenough-Harris Drawing Test with a sample of 23 children. A correlation of .60 was obtained between the Binet IV and the PPVT-R, which supports the use of the PPVT-R as a screening instrument. In addition, the PPVT-R and the Vocabulary

subtest of the Binet IV achieved a correlation of .53, suggesting that the two measures are tapping similar skills. The Binet IV and Columbia Mental Maturity Scale received a lower correlation, .47, than did the Binet IV and the PPVT-R.

Finally, the relationship between the PPVT-R and the Stanford-Binet IV was evaluated with 32 college students (Carvajal, Gerber & Smith, 1987). Results revealed a statistically significant correlation of .69 between the Stanford-Binet composite Standard Age Score (SAS) and the PPVT-R Standard Score Equivalent (SSE), which is comparable with the median value of .62 reported for 72 correlational studies dealing with the PPVT and the 1960 Stanford-Binet (Dunn & Dunn, 1981). The findings from this study, as well as from previously mentioned studies by Carvajal et al. (1987), suggest that the PPVT-R may be the instrument of choice for screening purposes.

Anastasi (1988) suggested that intelligence tests correlate about as highly with achievement tests as various intelligence tests correlate with each other. Breen (1983) focused on the comparability between the PPVT-R and three clusters of the Woodcock-Johnson Psycho-Educational Test Battery with 28 regular education (RE) and 28 learning disabled (LD) students.

The data suggested that the PPVT-R retained limited measurement capacity of academic achievement as compared with the Woodcock-Johnson.

The stability and equivalence of forms of the PPVT-R was assessed by Bracken and Prasse (1983); Argulewicz, Bingenheimer & Anderson (1983); Worthing, Phye & Nunn (1984). These studies obtained appropriate equivalence of forms with Anglo-American, Mexican-American, "at risk", handicapped, and learning disabled students.

Development of the McCarthy Scales of Children's Abilities

The McCarthy Scales of Children's Abilities (MSCA) (McCarthy, 1972) were developed in order to satisfy the need for a single instrument to assess a child's general intellectual level. The MSCA was designed for evaluating children aged 2 years, 6 months to 8 years, 6 months. Special instruction and training are needed to administer the test, which typically requires 45 to 50 minutes for children 2 years, 6 months to 5 years and 60 to 75 minutes for children 5 years to 8 years, 6 months. The MSCA have received praise for the game-like and nonthreatening nature of the tasks incorporated in the test as well as for the variety of tasks that enable the examiner to retain the child's

interest.

The McCarthy Scales are comprised of 18 subtests or tasks which include: Block Building, Puzzle Solving, Pictorial Memory, Word Knowledge, Number Questions, Tapping Sequence, Verbal Memory, Right-Left Orientation, Leg Coordination, Arm Coordination, Imitative Action, Draw-a-Design, Draw-a-Child, Numerical Memory, Word Fluency, Counting and Sorting, Opposite Analogies and Conceptual Grouping. The 18 previously mentioned subtests collectively form the 6 subscales of the MSCA: Verbal, Perceptual-Performance, Quantitative, General Cognitive, Memory and Motor. The Verbal Scale (V) examines the child's ability to express himself verbally and determines the maturity level of his verbal concepts. The Perceptual-Performance Scale (P) evaluates a child's reasoning ability through tasks that do not require the child to speak. The Quantitative Scale (Q) assesses the child's manageability of numbers and understanding of quantitative words. The Memory Scale (Mem) examines a child's short-term memory through verbal and nonverbal responses. The Motor Scale (Mot) tests a child's coordination as one performs various gross and fine motor tasks. Each of the previously mentioned scales have a mean of 50 and a standard deviation of 10.

The General Cognitive Scale (GCI) encompasses all of the tests in the Verbal, Perceptual-Performance and Quantitative Scales, which provides a measure of the child's overall cognitive abilities. Only 3 of the 18 subtests in the MSCA are excluded from the General Cognitive Scale: Leg Coordination, Arm Coordination and Imitative Action. The GCI has a mean of 100 and a standard deviation of 16, which allows the McCarthy Scales to be comparable to other intelligence tests. However, McCarthy does not utilize the term IQ due to the misinterpretations and unfortunate connotations associated with the term. The GCI was intended to indicate a child's cognitive level in relation to other children of the similar chronological age. For each of the 6 scales, the child's raw score is converted into a scaled score, an index. The GCI is presented as an index of a child's functioning level at a particular point in time.

Review of the Literature on the McCarthy Scales of Children's Abilities

Studies investigating the reliability and validity of the MSCA have been conducted since 1972; however, only sporadic attention has been given to this instrument in the literature (Bryant & Roffe, 1978). For the General Cognitive Scale, the average

reliability coefficient for the ten age groups is .93. The average correlation coefficients for the other scales ranged from .79 to .88, while the only reliability coefficient below .70 was obtained on the (Mot) at ages 6 years 6 months to 8 years, 6 months (McCarthy, 1972).

In 1973, Kaufman analyzed the MSCA in terms of Guilford's structure of intellect model in order to determine similar features. Results revealed analogous measures between Guilford's model and the MSCA. In addition, the study evaluated the abilities assessed by the MSCA, Stanford-Binet, Wechsler Preschool and Primary Scale of Intelligence (WPPSI) and Wechsler Intelligence Scale of Children (WISC), which revealed that the MSCA and Stanford-Binet measure similar percentages of the structure of intellect operations, contents, and products.

Kaufman and Hollenbeck (1973) factor analyzed the standardization edition of the MSCA. The authors analyzed 3 age groups in the study for a sum of 373 subjects. The means revealed a marked progression with age, which is an indicator of the age-relatedness of the tasks. Additionally, the majority of the tasks were of an appropriate difficulty level and at each of these levels, 3 consistent factors emerged: General

Cognitive, Memory and Motor. These consistent factors for each age level supported the stability of the underlying structure of the MSCA.

The stability of the McCarthy Scales was evaluated by Davis and Slettedahl (1976), implementing a test-retest interval of 1-year. The sample consisted of 43 children of a mixed culture with a median age of 5 years, 8 months. Test-retest correlations following the one-year period were in the .60s to .80s.

Bryant and Roffe (1978) examined the internal consistency of the MSCA with 38 middle class subjects ranging in age from 5 years, 5 months to 6 years, 5 months. The test-retest interval utilized for this study ranged from 3 to 6 weeks. Intercorrelations for the 5 scales were high. Results indicated that the MSCA is a relatively stable assessment instrument, allowing it to be competitive with other assessment instruments.

A moderate amount of research has been conducted on the construct validity of the MSCA. Analysis of the standardization data has yielded favorable support for the construct validity of the battery for normal children (Kaufman & Kaufman, 1975). The concurrent validity of the MSCA was examined by implementing four frequently utilized tests: the Stanford-Binet,

Developmental Test of Visual-Motor Integration, Goodenough-Harris Drawing Test and the Denver Developmental Screening Test, all of which assess cognitive, motor and perceptual development. The sample consisted of 46 preschool children ranging in age from 2 years, 8 months to 5 years, 1 month. The highest correlation between instruments was obtained between the MSCA GCI and the Stanford-Binet IQ. Lower but significant correlations were found for the MSCA GCI and the Developmental Test of Visual-Motor Integration and .48 for the MSCA GCI and the Goodenough-Harris Drawing Test. The authors concluded that the McCarthy Scales and the Stanford-Binet measure similar abilities with approximately the same accuracy (Krohn & Traxler, 1979).

Arinoldo (1982) reviewed the concurrent validity of the MSCA with the WPPSI and WISC-R. The subjects were 40 children who were administered the MSCA and the age-appropriate Wechsler scale. Significant moderate to strong correlations were noted between the MSCA GCI and the Wechsler IQs for this population.

Lastly, the relationship between the scores on Stanford-Binet IV and the MSCA were examined in a 1988 study. The sample included 21 children from a kindergarten class. The subjects were given the

complete battery of the Binet IV as well as the general cognitive area of the MSCA. A correlation of .68 was obtained between the Binet IV and the MSCA composite score. The results of this study would suggest that both tests measure similar constructs (Carvajal, Karr, Hardy & Palmer, 1988).

Studies were conducted with the McCarthy Scales to evaluate such variables as race, gender and social class differences. In 1973, Kaufman and Kaufman found black and white children did not differ significantly on any of the cognitive scales between the ages of 2 years, 6 months to 5 years, 6 months. However, a significant difference was obtained in which white children scored about one-half standard deviation higher than the black children at ages 6 years, 6 months through 8 years, 6 months. In addition, the black children scored significantly higher on the Motor Scale at ages 4 years to 5 years, 6 months. Overall, the relatively few significant differences that were discovered between white and black children on the MSCA should enhance the utilization of this instrument for all children, regardless of race.

A factor analysis of both black and white children's scores was conducted by Kaufman and Dicuio in 1975. The subjects were 124 black children and 688

white children from the standardization sample who were evaluated in order to determine the construct validity of the instrument for each racial group. As a result of the factor analysis, it was determined that the factors which emerged for both groups of children were congruent.

In 1973, Kaufman and Kaufman investigated gender differences on the McCarthy Scales. The sample included 50 boys and 50 girls from the standardization sample at each of the 10 age levels between 2 years, 6 months and 8 years, 6 months. The girls achieved higher mean GCIs than the boys at all seven age levels; however, none of the differences reached significance. This study supported the utilization of one set of norms for both genders.

Kaufman and Kaufman (1975) examined the relationship of social class to the cognitive and motor indexes of the MSCA with a sample of 154 black children and 862 white children. The subjects were divided into five occupational categories according to their father's occupation in order to determine social-class differences. Children categorized as middle-class scored significantly higher for both racial groups than did working class youngsters on each of the six indexes.

Kaufman pursued the possibility of developing a short form of the MSCA in order to reduce administration and interpretation time of the test. Two short forms of the MSCA, Kaufman's McCarthy Short Form (1977) and the McCarthy Screening Test (1978), were developed. The Kaufman's McCarthy Short Form includes: Puzzle Solving, Word Knowledge, Numerical Memory, Verbal Fluency, Counting and Sorting and Conceptual Grouping, while the McCarthy Screening Test contains the following six subtests: Right-Left Orientation, Draw-a-Design, Numerical Memory, Verbal Memory, Leg Coordination and Conceptual Grouping.

An additional short form was developed by Taylor, Slocumb and O'Neill (1979). The authors examined the MSCA in order to construct a short form that would best predict the GCI. The MSCA was administered to 50 kindergarten-aged children to obtain general cognitive indices. A stepwise regression analysis was incorporated to select the 6 subtests that best predicted the GCI. The following subtests were selected: Counting and Sorting, Pictorial Memory, Number Questions, Verbal Fluency, Numerical Memory, and Tapping Sequence. A correlation of .96 was obtained between the subjects' GCIs and their performance on the previously mentioned subtests.

Harrison and Naglieri (1981) reviewed the predictive validities of the 2 McCarthy short forms with 53 first graders in Georgia. The Metropolitan Achievement Test was administered to the students nine months following the administration of the McCarthy short forms for comparison purposes. A correlation of .71 was found between Kaufman's estimated General Cognitive Index and the Metropolitan Basic Battery raw score. In addition, biserial correlation between "at risk" and "not at risk" classifications of the McCarthy Screening Test and the Metropolitan Basic Battery raw scores ranged from .43 to .78. It would appear from this study that Kaufman's short form incorporates better predictive validity than the McCarthy Screening Test for this population.

The factor structure of the McCarthy Screening Test was investigated with a sample of 555 children whose mean age was 61.4 months. All of the test correlations for the 1-factor solutions (GCI) were moderately high, which implies that the tests are measuring components of the same attribute (cognitive and sensorimotor functions), but to somewhat differing degrees (Vance, Blixt & Kitson, 1982).

Review of the Literature on the Peabody Picture Vocabulary Test-Revised and the McCarthy Scales of Children's Abilities

The remainder of the review will deal with studies comparing the tests that will be used in this study. Limited research has been conducted on the PPVT-R and the MSCA exclusively; therefore, any new knowledge that could be contributed to this area would be beneficial.

The relationships between the PPVT-R, MSCA and the Peabody Individual Achievement Test (PIAT) were explored with a sample of 26 children who were randomly selected from three elementary classes. The PPVT-R correlated significantly with the PIAT total score, as well as with the Verbal, Quantitative, Memory, and GCI of the MSCA. In addition, the mean PPVT-R score and the MCSA GCI were nearly identical, 104.6 and 104.4, respectively. The author suggested that the PPVT-R and the MSCA should not be considered interchangeable despite the significant correlation that was found between the PPVT-R and the Verbal sections of the MSCA (Naglieri, 1981).

In 1982, Gullo and McLoughlin examined the scores of normal preschool children on the PPVT-R and the MSCA. The sample consisted of 30 children between the ages of 3 and 4 years of age who were randomly selected

from four nursery school classes. Correlations were computed between the PPVT-R and the MSCA for the total group of 30 children. A significant correlation was obtained between the PPVT-R and the MSCA GCI, as well as with the Verbal, Perceptual-Performance, Quantitative, and Motor Scale. These results indicate that the PPVT-R measures the overall cognitive ability of preschoolers in a similar manner as the MSCA GCI. When comparing the mean scores of the 2 instruments, Gullo and McLoughlin determined that the PPVT-R standard score was equivalent to the MSCA GCI for the 4-year-olds, while it underestimated the GCI by approximately 13 points for the 3-year-olds. The authors suggested that this difference may reflect skills assessed by the MSCA at different age levels, but warrants further investigation.

The final investigation compared the scores between the PPVT-R and the MSCA for 35 preschool children who were designated as "at risk" children. Significant, but moderate correlations were obtained between the PPVT-R and the MSCA GCI. Whereas Naglieri (1981) judged the highest correlations between the PPVT-R and the MSCA to exist between the MSCA's verbally loaded subtests, this study determined moderate correlations between the PPVT-R and the MSCA

throughout the scales. The moderate correlations established between the two instruments imply that somewhat different skills are being assessed (Bracken & Prasse, 1983).

The literature reviewed above has demonstrated favorable qualities of both instruments examined in this study. The research has suggested that the PPVT-R is a valid instrument which measures similar qualities on the PPVT and other intelligence tests. In addition, in the majority of comparison studies, the publishers' contention of equivalency of forms has been substantiated. However, when PPVT-R standard scores are compared to PPVT IQs, WISC-R IQs and Stanford-Binet IQs, the mean PPVT-R scores have been significantly lower (Altepeter & Handal, 1985).

The McCarthy has the advantage of assessing a variety of abilities, which yields a vast amount of meaningful educational data. The MSCA display major strengths that identify the instrument as one of the best broad-based diagnostic tools for the assessment of children's abilities. Despite its favorable qualities, the McCarthy has received criticism due to the exclusion of exceptional children from the standardization sample. In general, the MSCA does not provide a high enough ceiling for the exceedingly

superior child of six, seven, or eight years. However, precise measurement of an exceptional child's deviation from the norm proves less essential than the determination of the child's strengths and abilities (Taylor, 1979).

Purpose of the study

The purpose of this study is to establish concurrent criterion-related validity of the PPVT-R by comparing it with the MSCA. The Peabody has received favorable reviews and is the most frequently utilized screening test available (Lubin, Larsen, & Matarazzo, 1984). The criterion of the MSCA was implemented as this instrument retains superior psychometric properties. Also, a limited amount of research has been conducted on these two instruments exclusively.

Statement of Significance

It would seem to be of significant clinical importance to determine the comparability between a single-skill test, the PPVT-R, and a multi-skill test, such as the MSCA. The results of this study would enable psychologists and educational diagnosticians who work with children to make additional judgments about the validity of these instruments, as well as to contribute knowledge to the field. While both are older assessment instruments, it is of utmost

importance to provide psychologists with options for testing or retesting a child's intellectual level.

CHAPTER 2

METHODS

Sample

This study included 32 children from a midwestern city of approximately 27,000. The children included in this study ranged in age from 4 years, 2 months to 6 years, 3 months. Permission to test children at Butcher School in Emporia, Kansas, was obtained from Kansas Unified School District 253 as was parental consent for the children's participation.

The parents of the participants were requested to read and sign an informed consent form prior to the data collection. The consent form contained information regarding the intent and purpose of the study, the methods implemented in the study, as well as the rights of the subjects. The parents' signature substantiated approval for the child to participate in the study. Additionally, in order to satisfy university regulations, an application for testing was submitted to the Emporia State University's Review Board for Treatment of Human Subjects prior to the data collection.

Procedure

The PPVT-R and the MSCA were individually administered to each subject. The tests were given to

the participants in a counterbalanced order, with the data being collected over a three month interval. To allow for accurate scoring, the age, gender and name of the subject were obtained.

The PPVT-R was individually administered by one of two qualified examiners in rooms appropriately designed for testing at Emporia State University. The MSCA was individually administered by a qualified examiner as well in a similar fashion as the PPVT-R. Both tests were administered and scored in strict accordance with the appropriate manual.

Statistical Design

This study yielded five scores per subject: four scores from the MSCA (Verbal Index, Perceptual-Performance Index, Quantitative Index, and General Cognitive Index) and one from the PPVT-R (Standard Score Equivalent). The scores were analyzed by calculating the Pearson product-moment correlation coefficient (Pearson *r*) to determine the association between the PPVT-R and MSCA scores. A 2 X 2 X 5 Mixed Factor Analysis of Variance was calculated to determine differences in gender (Boys, Girls), grade-level (Preschool, Kindergarten), and test scores (Verbal, Perceptual-Performance, Quantitative, General Cognitive Index, Peabody Picture Vocabulary Test-

Revised Standard Score Equivalent). The means and standard deviations were also analyzed by comparing the findings with the normative group.

CHAPTER 3

RESULTS

Four scores from the MSCA and one score from the PPVT-R were obtained for 32 children (17 boys and 15 girls). The children ranged in age from 4 years, 2 months to 6 years, 3 months. The means and standard deviations of the test scores are presented in Table 1. All of the mean scores reported in Table 1 were somewhat higher than the normative group scores.

Table 1

Means and Standard Deviations for the MSCA and PPVT-R

Test	M	SD
MSCA		
Verbal	55.03	7.08
Perceptual-Performance	57.21	8.74
Quantitative	56.84	6.87
General Cognitive	111.84	10.67
PPVT-R		
Standard Score Equivalent	111.38	12.77

Tables 2 and 3 list the means and standard deviations of the test scores by gender. Table 2 presents the boys' results, while Table 3 presents the girls' results.

Table 2

Test Means and Standard Deviations for Boys

Test	M	SD
MSCA		
Verbal	56.71	7.17
Perceptual-Performance	55.59	6.62
Quantitative	56.12	6.72
General Cognitive	111.53	10.79
PPVT-R		
Standard Score Equivalent	114.06	14.25

Table 3

Test Means and Standard Deviations for Girls

Test	M	SD
MSCA		
Verbal	53.13	6.70
Perceptual-Performance	59.10	10.58
Quantitative	57.67	7.18
General Cognitive	112.20	10.90
PPVT-R		
Standard Score Equivalent	108.33	10.51

The mean scores reported in Tables 2 and 3 are above the normative samples for the MSCA and PPVT-R. However, none of these differences were significant. The boys scored higher than girls in regard to mean scores on the Verbal Scale of the MSCA and the PPVT-R Standard Score Equivalent. The girls scored higher than the boys in regards to the mean scores on the Perceptual-Performance Scale, Quantitative Scale and General Cognitive Index.

A 2 X 2 X 5 mixed factor analysis of variance was conducted to examine differences in gender (Boys, Girls), grade-level (Preschool, Kindergarten) and test

scores (Verbal, Perceptual-Performance, Quantitative, General Cognitive Index and Peabody Picture Vocabulary Test-Revised Standard Score Equivalents). The raw scores were converted to standard z scores for the analysis. No significant main effects were obtained for gender, grade-level or tests, nor were any of the interactions significant.

Table 4 presents the results of the ANOVA.

Table 4

ANOVA Summary Table for Tests by Gender and Grade

Source	df	SS	MS	F	p
Between					
Gender	1	0.02	0.02	0.01	ns
Grade	1	3.26	3.26	2.52	ns
Gender X Grade	1	0.19	0.19	0.15	ns
Error	28	36.26	1.29		
Within					
Tests	4	1.58	0.40	1.04	ns
Tests X Gender	4	3.52	0.88	2.31	ns
Tests X Grade	4	2.04	0.51	1.34	ns
Tests X Gen. X Gr.	4	2.64	0.66	1.74	ns
Error	112	42.64	0.38		

Pearson product-moment correlation coefficients were calculated between the MSCA scores and the PPVT-R. Although the means of the total test scores of the two tests were virtually identical, the MSCA GCI and the PPVT-R SSE correlated poorly (r = .16, p<.05). The PPVT-R also failed to correlate significantly with the Verbal, Perceptual-Performance and Quantitative Scales (.34, .13, and .12, respectively, p<.05).

CHAPTER 4

DISCUSSION

It is essential that individuals involved in the assessment of children's intelligence have information on the validity of testing instruments. This information can be utilized in the selection of proper instruments for gaining knowledge about a child's ability level. Accuracy of measurement is vital because a child's performance on an intelligence measure often has a tremendous impact on future academic placement. If erroneous conclusions are drawn from a test that was improperly used, a child's academic progress could be hindered. Therefore, diagnosticians need to be aware of the specific abilities that a test is assessing as well as the uses and limitations of the tests. For example, if one is implementing a single-skill test for assessment, caution is warranted in generalizing these findings to areas which require a multi-skill test of intelligence.

Prior to a test gaining acceptance as an adequate instrument for measuring intelligence, its concurrent validity must be substantiated. The present study was designed to assess the concurrent criterion-related validity of the PPVT-R by comparing it with the MSCA.

The results of the present study revealed that the

two instruments produced almost identical mean scores. In addition, only a moderate difference existed between genders on the Verbal, Perceptual-Performance, Quantitative and General Cognitive Index. However, the males scored higher than the females on the PPVT-R SSE, but not to a significant degree.

Although the overall mean scores on the two instruments were very similar, the tests correlated at a low level. The similar mean scores would indicate that the two instruments have the potential of producing generally comparable results. Nonetheless, the low correlation coefficient obtained in this study suggests enough variability that one should not assume a direct relationship between scores on the two tests.

The evidence that supports the use of the PPVT-R as a measure of general intelligence is inconsistent. Therefore, the PPVT-R should not be substituted for a more comprehensive measure of intelligence, but rather serve as an appropriate screening device. The literature supports the use of the PPVT-R as an effective screening measure in determining the need for a more broad-based assessment.

Two important factors may have influenced the outcome of this study which should be considered for future research. The sample included in this study was

limited in size and diversity. The children in this study were selected from one institution and may not be truly representative of the entire population of this age range. Therefore, a larger number of individuals from a more diverse population would enhance the sample implemented in similar studies.

REFERENCES

Altepeter, T.S. (1989). The PPVT-R as a measure of psycholinguistic functioning: A caution. Journal of Clinical Psychology, 45, 935-940.

Altepeter, T.S., & Handal, P.J. (1985). A factor analytic investigation of the use of the PPVT-R as a measure of general achievement. Journal of Clinical Psychology, 41, 540-543.

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

Argulewicz, E.N., Bingenheimer, L.T., & Anderson, C.C. (1983). Concurrent validity of the PPVT-R for Anglo-American and Mexican-American students. Journal of Psychoeducational Assessment, 1, 163-167.

Arinoldo, C.G. (1982). Concurrent validity of the McCarthy Scales. Perceptual and Motor Skills, 54, 1343-1346.

Beck, F.W., & Black, F.L. (1986). Comparison of PPVT-R and WISC-R in a mild/moderate handicapped sample. Perceptual and Motor Skills, 62, 891-894.

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.

Breen, M.J. (1981). Comparison of the Wechsler Intelligence Scale for Children-Revised and the Peabody Picture Vocabulary Test-Revised for a referred population. Psychological Reports, 49, 717-718.

Breen, M.J. (1983). A correlational analysis between the PPVT-R and Woodcock-Johnson achievement cluster scores for nonreferred regular education and learning disabled students. Psychology in the Schools, 20, 295-297.

Breen, M.J., & Siewart, J.C. (1983). Comparison of the Peabody Picture Vocabulary Test-Revised and the Wechsler Intelligence Scale for Children-Revised for learning-disabled and referred students. Journal of Psychoeducational Assessment, 1, 95-99.

Bryant, C.K. & Roffe, M.W. (1978). A reliability study of the McCarthy Scales of Children's Abilities. Journal of Clinical Psychology, 34, 401-406.

Candler, A.C., Maddux, C.D., & Johnson, D.L. (1986). Relationship of scores on PPVT-R and WISC-R with special education children and youth. Perceptual and Motor Skills, 62, 417-418.

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., Karr, S.K., Hardy, K.M., & Palmer, B.L. (1988). Relationships between scores on Stanford-Binet IV and scores on McCarthy Scales of Children's Abilities. Bulletin of the Psychonomic Society, 26, 349.

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, 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.

Davis, E.E., & Slettedahl, R.W. (1976). Stability of the McCarthy Scales over a one-year period. Journal of Clinical Psychology, 32, 798-800.

Davis, S.E., & Kramer, J.J. (1985). Comparison of the PPVT-R and WISC-R: A validation study with second-grade students. Psychology in the Schools, 22, 265-267.

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

Ferrari, M. (1980). Comparisons of the Peabody Picture Vocabulary Test and the McCarthy Scales of Children's Abilities with a sample of autistic children. Psychology in the Schools, 17, 466-469.

Gullo, D.F., & McLoughlin, C.S. (1982). Comparison of scores for normal preschool children on Peabody Picture Vocabulary Test-Revised and McCarthy Scales of Children's Abilities. Psychological Reports, 51, 623-626.

Halpin, G., Simpson, R.G., & Martin, S.L. (1990). An investigation of racial bias in the Peabody Picture Vocabulary Test-Revised. Educational and Psychological Measurement, 50, 183-188.

Harrison, P.L., & Naglieri, J.A. (1981). Comparison of the predictive validities of two McCarthy short forms. Psychology in the Schools, 18, 389-393.

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.

Hollinger, C.L., & Sarvis, P.H. (1984). A comparison of the PPVT-R and WISC-R with rural children referred for assessment. Psychology in the Schools, 21, 97-101.

Kaufman, A.S. (1973). Analysis of the McCarthy Scales in terms of Guilford's structure of intellect model. Perceptual and Motor Skills, 36, 967-976.

Kaufman, A.S., & Dicuio, R.F. (1975). Separate factor analyses of the McCarthy Scales for groups of black and white children. Journal of School Psychology, 13, 10-17.

Kaufman, A.S., & Hollenbeck, G.P. (1973). Factor analysis of the standardization edition of the McCarthy Scales. Journal of Clinical Psychology, 29, 358-362.

Kaufman, A.S., & Kaufman, N.L. (1973). Black-white differences at ages 2 1/2 - 8 1/2 on the McCarthy Scales of Children's Abilities. Journal of School Psychology, 11, 196-206.

Kaufman, A.S., & Kaufman, N.L. (1973). Sex differences on the McCarthy Scales of Children's Abilities. Journal of Clinical Psychology, 29, 362-365.

Kaufman, A.S., & Kaufman, N.L. (1975). Social-class differences on the McCarthy Scales for black and white children. Perceptual and Motor Skills, 41, 205-206.

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

Krohn, E.J., & Traxler, A.J. (1979). Relationship of the McCarthy Scales of Children's Abilities to other measures of preschool cognitive, motor and perceptual development. Perceptual and Motor Skills, 49, 783-790.

Kutsick, K., Vance, B., Schwarting, F.G., & West, R. (1988). A comparison of three different measures of intelligence with preschool children identified at-risk. Psychology in the Schools, 25, 270-274.

Levenson, R.L., & Lasher-Adelman, V. (1988). Stability of the Peabody Picture Vocabulary Test-Revised for emotionally handicapped middle-school-aged children. Perceptual and Motor Skills, 67, 392-394.

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.

McCarthy, D. (1972). McCarthy Scales of Children's Abilities manual. New York, NY: The Psychological Corporation.

Naglieri, J.A. (1981). Concurrent validity of the revised Peabody Picture Vocabulary Test. Psychology in the Schools, 18, 286-289.

Prasse, D.P., & Bracken, B.A. (1981). Comparison of the PPVT-R and WISC-R with urban educable mentally retarded students. Psychology in the Schools, 18, 174-177.

Prout, H.T., & Schwartz, J.F. (1984). Validity of the Peabody Picture Vocabulary Test-Revised with mentally retarded adults. Journal of Clinical Psychology, 40, 584-587.

Rosso, M., Falasco, S.L., & Koller, J.R. (1984). Investigations into the relationship of the PPVT-R and the WISC-R with incarcerated delinquents. Journal of Clinical Psychology, 40, 588-591.

Smith, T.C., Edmonds, J.E., & Smith, B. (1989). The role of sex differences in the referral process as measured by the Peabody Picture Vocabulary Test-Revised and the Wechsler Intelligence Scale for Children-Revised. Psychology in the Schools, 26, 354-358.

Sommers, R.K., Erdige, S., & Peterson, M.K. (1978). How valid are children's language tests. Journal of Special Education, 12, 393-406.

Taylor, R.L. (1979). Comparisons of the McCarthy Scales of Children's Abilities and the Peabody Picture Vocabulary Test. Psychological Reports, 45, 196-198.

Taylor, R.L., Slocumb, P.R., & O'Neill, J. (1979). A short form of the McCarthy Scales of Children's Abilities: Methodological and clinical applications. Psychology in the Schools, 16, 347-350.

Vance, B., Blixt, S., & Kitson, D.L. (1982). Factor structure of the McCarthy Screening Test. Psychology in the Schools, 19, 33-38.

Vance, B., Kitson, D., & Singer, M. (1983). Further investigation of comparability of the WISC-R and PPVT-R for children and youth referred for psychological services. Psychology in the Schools, 20, 307-310.

Vance, B., West, R., & Kutsick, K. (1989). Predictions of Wechsler Preschool and Primary Scale of Intelligence IQ scores for preschool children using the Peabody Picture Vocabulary Test-Revised and the Expressive One Word Picture Vocabulary Test. Journal of Clinical Psychology, 45, 642-644.

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.

Wright, D. (1983). Effectiveness of the PPVT-R for screening gifted students. Psychology in the Schools, 20, 25-26.

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

FROM: Emporia State University Graduate School

I, Gina L. Lomax, hereby submit this thesis/report to Emporia State University as partial fulfillment of the requirements for an advanced degree. I agree that the library of the university may make it available for use in accordance with its regulations governing materials of this type. I further agree that quoting, photocopying, or other reproduction of this document is allowed for private study, scholarship (including teaching) and research purposes of a nonprofit nature. No copying which involves potential financial gain will be allowed without written permission of the author.

Gina L. Lomax
Signature of Author

April 22, 1992
Date

Relationships Between Scores on the Peabody Picture Vocabulary Test-Revised and the McCarthy Scales of Children's Abilities

Title of Thesis/Research Project

Signature of Graduate Office Staff Member

April 22, 1992
Date Received

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