AN ABSTRACT OF THE THESIS OF

Charlene A. Ziegler for the Master of Science

in Psychology presented on December, 1991

Title: Relationship Between Scores on the Columbia Mental Maturity Scale and the Peabody Picture Vocabulary Test-Revised

Abstract Approved: David L. Vaughn

Screening instruments are vitally important for determining whether or not a child may need a further and more complete assessment. They can also help school personnel make decisions concerning the selection of curriculum materials and learning tasks that are available for all the children. Screening instruments are used to make such decisions because they are quick to administer and cost effective.

The present study was designed to establish concurrent criterion-related validity. The relationship between the Peabody Picture Vocabulary Test-Revised (PPVT-R) and the Columbia Mental Maturity Scale (CMMS) was investigated. Fifty-one children (27 boys and 24 girls) were tested with both instruments. They ranged in age from 3 years to 7 years.

The CMMS produces one score referred to as the Age Deviation Score (ADS) and the PPVT-R produces one score referred to as a Standard Score Equivalent (SSE). The mean scores obtained in this study for both tests were above those of the normative samples. The females scored higher

than the males in regards to mean scores on the CMMS, and the males scored higher than the females in regards to mean scores on the PPVT-R. Furthermore, a Pearson product-moment coefficient was calculated to determine the relationship between the CMMS and PPVT-R. It was very low ($r = .18$).

A 2 X 2 factorial design was conducted to determine if there were gender differences between the CMMS and the PPVT-R scores. The Gender X Tests was significant, $F(1,49) = 6.26$, $p<.05$. The significant differences were found between the males and females scores on the PPVT-R. The males scored significantly higher. Also, the males scored significantly higher on the PPVT-R than they scored on the CMMS ($p<.05$). The similar mean scores suggest that the two tests are capable of producing comparable assessments. Nevertheless, the low correlation coefficient demonstrates that the two tests do not produce comparisions of the same equality.

Relationship Between Scores on the Columbia Mental Maturity Scale 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

Charlene Ziegler

December, 1991

Stephen F. Davis
Approved for the Major Division

Faye N. Vowell
Approved for the Graduate Council

ACKNOWLEDGEMENTS

I would like to thank Dr. David Dungan for his guidance throughout this study. I would also like to extend my sincere appreciation to Mr. Howard H. Carvajal for his assistance and continuous support during this study. I would also like to thank Dr. Phillip Wurtz for serving on my thesis committee, and Dr. Kenneth Weaver for assisting me with the statistical analysis.

I want to thank David Anderson for all the love, patience, and encouragement he has given me throughout graduate school. Finally, I want to express thanks from the bottom of my heart to my parents Betty Jean Ziegler and the late Lawrence Max Ziegler. I thank my mother for all the love, support, and assistance she has given me throughout my educational journey. I also thank my father for his love and support while he was with us, and since his death I know he has helped me in his own special way. I dedicate this thesis to my wonderful, loving parents, for without them I would not have made it this far.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

CHAPTER 1

Introduction

A special interest in intellectual assessment developed during the nineteenth century as a result of concern for the proper treatment of the mentally retarded and the insane. Because of this concern, special institutions were developed for each of these populations; thus, it became necessary to distinguish between the insane and the mentally retarded. According to Anastasi (1988), the first explicit statement of this distinction was published by Esquirol in 1838. In his two-volume work over one hundred pages were devoted to mental retardation. He pointed out that there were many degrees of retardation, varying from normality to "low-grade idiocy." To develop some system for classifying the degrees of retardation, Esquirol tried many procedures but found that an individual's use of language was the best criterion for determining one's intellectual level.

The first experimental psychologists studied human behavior in the laboratory concentrating on sensitivity to visual, auditory, and other sensory stimuli, and with reaction time (Anastasi, 1988). Many approaches were tried to assess intellectual abilities, but none proved to be very successful. Then, later on according to Anastasi (1988), Alfred Binet felt that children who did not respond to normal schooling should be examined before being dismissed, and if they were considered educable, they should be assigned to special classes. Binet and the members of the Society for the Psychological Study of the Child urged the Ministry of Public Instruction to improve the education conditions of retarded children. In 1904, the Minister of Public Instruction appointed Binet to a committee to study procedures for the education of retarded children. In association with the objectives of the committee, in 1905 Binet with the help of Simon, prepared the first Binet-Simon Scale. This was the first practical intelligence test; success was achieved. Many of the concepts and processes of the Binet-Simon scale formed the model for present day intelligence tests.

For the past 80 years, a vast amount of research has been conducted, and numerous tests have been published, as well as many revisions of old tests in order to evaluate individuals' intellectual functioning. Because of this constant stream of new tests, psychologists, as well as other test givers and interpreters, need to be skilled and informed about the tests they are using. It is important to know how reliable a test is, how valid it is, and how it was standardized. With each test, there should be a manual to provide essential and critical information about the test, such as its reliability and validity.

Today, due to lack of time and money, the individuals involved in testing are always looking for a quick and cost effective method for accurately assessing children's abilities. To meet these needs two tests are commonly used, the Columbia Mental Maturity Scale (CMMS) and the Peabody Picture Vocabulary Test-Revised (PPVT-R). These quick tests are usually referred to as screening instruments. In the context of this study, screening is a process of measuring the intelligence of infants and children to identify those needing further and more complete assessment to determine whether they are at risk, or if they may be at risk in the future (Bailey & Wolery, 1989). Screening programs have also been developed in response to federal legislation. "Federal legislation (in the form of PL 94-142, the Education for the Handicapped Act, and PL 99-457, the 1986 Amendments to PL 94-142) mandates that states develop and implement public awareness programs focusing on early identification of handicapped infants and toddlers and a comprehensive 'child find' system for referring and screening children to identify those who should participate in a more extensive evaluation" (Bailey & Wolery, 1989, p.2). It is important to identify children who need further evaluation early so that an Individual Educational Plan can be developed, and the child's needs can be met. Also psychologists and school personnel must periodically make decisions concerning the selection of curriculum materials and the learning tasks they have set forth for all children.

The results obtained from testing can be very helpful when making such decisions.

Because there are numerous intelligence tests, and they differ in content, the scores yielded from every test need to be interpreted carefully. It is important not to overgeneralize the score from a single-skilled test. If further assessment is needed, a multi-skilled test must be used. Most screening tests are single-skilled tests. The CMMS and the PPVT-R are well known and widely used tests of this type. In fact, according to a national survey of psychological test usage (Lubin, Larsen, Matarazzo, & Seever, 1985), the PPVT ranked eleventh, out of the thirty most frequently used tests. Furthermore, Kaufman (1978) indicated that "The CMMS is undoubtedly the best brief instrument (verbal or nonverbal) available, and it ranks as one of the finest test for assessing preschool children" (p. 301). The value of intelligence tests is usually proven by comparing the scores to a well known test, such as the Wechsler Scales or the Stanford-Binet. Both the CMMS and the PPVT-R have been compared to these instruments. Since both the CMMS and the PPVT-R are very popular and are used frequently, psychologists and all individuals who administer such tests will want to know what the relationship is between these two tests. These screening instruments guide the examiner in determining if a more comprehensive assessment is needed.

Review of Literature

Columbia Mental Maturity Scale

The first edition of the CMMS was published in 1954, the second in 1959, and the third in 1972. Burgemeister, Blum, and Lorge (1972) suggest that the first edition of the CMMS was developed to provide a satisfactory instrument for estimating the mental capacity of handicapped children, particularly at the very early ages. This type of test demands no verbal response and very little motor response. The selection of items for the first CMMS was done with the intent of obtaining a measure of general mental ability for those children evaluated. However, according to Hiskey (1965), "The second edition (1959) was a revision of the original scale (1954), that took place after reports of research gave highly conflicting evidence in regard to validity and reliability coefficients, adequacy of norms, item difficulty, and the rationale for responses" (p. 800).

After extensive research on the second edition, a third edition was devised to replace the second, and the third edition is still being used today. The CMMS is an individually administered nonverbal measure of "general reasoning ability" (Egeland, 1965, p. 298). According to Burgemeister et al. (1972) reasoning includes both perceptual classification and higher-level abstract manipulation of symbolic concepts.

The CMMS was designed for children aged 3 years, 6 months through 9 years, 11 months. The test consists of 92, 6 x 19 inch cards, each consisting of three to five drawings. These cards are arranged in a series of eight overlapping levels, and each level contains anywhere from 51 to 65 items. The appropriate level is determined by the subject's chronological age. One card at a time is presented to the child, and he/she is to select the drawing which does not belong with the others by pointing to it. In all, the CMMS takes approximately 15 to 20 minutes to administer. The standard score obtained is called the Age Deviation Score (ADS), indicating the subject's deviation from the average score for children of a specified age group. The ADS ranges from a low of 50 to a high of 150, with the mean set equal to 100 and the standard deviation being 16. There is also a table for converting the subject's raw score to a Maturity Index (MI). This designates the standardization age group most similar to the subject in terms of test performance (Burgemeister et al., 1972).

The CMMS (1972) was standardized with a sample of 2600 children aged 3 1/2 years to 9 years, 11 months who were stratified on the basis of geographic region, race, parental occupation, sex, and age. The geographic region was divided into four areas: Northeast, South, North Central, and West. The ethnic representation was divided into three groups: white, black, and other. Parental occupation was divided into six groups: professional, technical and managerial workers, clerical and sales workers, blue-collar workers, service workers, farm workers, and unemployed welfare recipients. There were 1300 boys and 1300 girls in the total of 2600 children. Finally, the age range was divided into 13 six-month age groups and 200 subjects within each age group were tested (Burgemeister et al., 1972).

The latest revision (1972) is a great improvement over the previous editions (1954 and 1959) because of the following important changes. To begin with, the items are grouped into eight levels which makes it easy to determine when testing should be terminated, which results in the expenditure of less testing time. Secondly, of the 92 items, 50 are new, two were unchanged, and the remaining 40 items were modified. Third, the ratio IQ was replaced by a standard score called Age Deviation Score (ADS), because of the difficulties with the use of the ratio IQ. Fourth, the new CMMS has a redesigned Individual Record Form, which was developed in order to be a better designed permanent individual test record by combining biographical data about the child with data about the test and the interpretation of derived scores. This new record form is suppose to communicate the nature and meaning of the CMMS to those who are unfamiliar with the test better than the previous record blanks. The final improvement is the standardization of the test, which has been discussed in the previous paragraph (Burgemeister et al., 1972; Egeland, 1965). In light of the reliability of the CMMS, according to the manual (Burgemeister et al., 1972) the split-half reliability coefficients ranged across ages from .85 to .91, and the test-retest reliability ranged from .84 to .86. These results were supported by Pascale (1973) when he tested 72 preschool children between the ages of three to five years, each with the Peabody Picture Vocabulary Test (PPVT), Raven's Coloured Progressive Matrices (CPM), and the CMMS (1972). The author then retested 36 children and arrived with a test-retest correlation coefficient of .85 for the CMMS, which was the highest of the three tests. Pascale also found that the correlation between the CMMS and the PPVT was .47. Although this was statistically significant, it accounts for less that 25 percent of the common variance.

In regard to the validity of the CMMS, the manual presents data from a group of 52 preschool and first-grade pupils who had been tested with the 1972 CMMS and the Stanford-Binet Intelligence Scale: Form L-M. The correlation between the two tests was .67. The manual also showed the correlation between the CMMS and the Otis-Lennon

Mental Ability Test. The data were reported for two different groups. One group consisted of 263 children and the other 90 children. The correlations were .69 and .62, respectively (Burgemeister et al., 1972).

<u>Peabody Picture Vocabulary Test-Revised</u>

The original PPVT was published in 1959 and the revised edition (PPVT-R) in 1981. The PPVT-R broadened or changed many features of the PPVT. For instance, the number of items for each test form was increased from 150 to 175, and about two-thirds of the stimulus words are new. Many items were redrawn to reflect a more appropriate racial, ethnic, and gender representation. Separate sets of test items have been provided for Forms L and M. Also, the PPVT IQ and Mental Age concepts were terminated, and now the raw scores derived from the PPVT-R may be converted to either deviation-typed age norms in the form of standard score equivalents, or to percentile ranks, stanines, or to developmental-age norms in the form of age equivalents. The standard score equivalents express in standard deviation units where the subjects score exceeds, or falls below the mean score of persons of the same age upon who the test was standardized. The mean of the PPVT-R is 100, and the standard deviation is 15. Also, prestandardization testing and calibration was done on 5,717 subjects to equate the level of difficulty of items in Forms L and M by using

Rasch-Wright latent traits analysis of items. Furthermore, the standardization of the PPVT-R was much improved from the PPVT (Wiig, 1985). According to McCallum (1985) the original PPVT was normed on a large but restricted sample of 4012 white individuals residing in and around Nashville, Tennessee. However, the Technical Manual (Robertson & Eisenberg, 1981) reports a much better standardization process for the PPVT-R. A stratified sampling procedure, based on population data from the 1970 U.S. Census, was used to establish quotas within each age group by sex, geographic region, occupation of major wage earner, race, and community size. The standardization process consisted of testing 4200 children and adolescents and 828 adults. Below age 19, the age range covered by the test was divided into nine 6-month age groups, from 2 years, 6 months to 6 years, 11 months, and twelve 1-year age groups, from 7 years through 18 years. One hundred females and 100 males were tested within each age group. The 19 through 40 year olds were stratified by age, sex, and occupation. The adults were broken into four age groups and just over 200 subjects were tested in each group. The United States was divided into four regions: Northeast, South, North Central, and West. Race was divided into four groups: white, black, Hispanic, and "other." Finally, children were selected from three types of communities: central cities, suburbs or small towns, and rural areas.

The PPVT-R provides an individually administered, norm-referenced test of receptive (hearing) vocabulary which includes a wide age range of 2 years, 6 months to 40 years (Wiig, 1985). Examinees are asked to indicate which of four pictures presented on a plate corresponds to a stimulus word read aloud by an examiner. This test is untimed, and takes about 15-20 minutes to administer (McCallum, 1985). There is a series of plates for Form L and for Form M, and each volume contains 175 test items preceded by 5 training plates; the plates are bound in an Easel-Book. For subject's under 8, training plates A, B, and C are given, and for subjects 8 and over, training plates D and E are given. For each form, there is a separate Individual Test Record. The record lists the stimulus words to be used with the training and test plates, and gives the key to the accurate choice. A space is provided for recording the subject's responses, raw score, standard score equivalent, percentile rank, and stanine scores with his/her error of measurement, and also for recording additional information about the subject (Dunn & Dunn, 1981).

According to the manual (Dunn & Dunn, 1981) the PPVT-R provides a quick estimate of one major aspect of verbal ability for those who have grown up in a standard English-speaking environment. The PPVT-R is not a comprehensive test of general intelligence; it measures only receptive vocabulary. It is useful for schools as an initial screening device in scanning for children who may need special attention. Clinically it is useful because it is less threatening due to the lack of verbal interaction, and it can be used successfully with certain autistic, withdrawn, and psychotic persons, as well as handicapped persons. The PPVT-R is also useful for vocational and research purposes.

When the authors of the PPVT-R, developed the item pool for tryout, they wanted twice as many items as the 350 planned for the final two test forms. After thoroughly reviewing the original PPVT, 144 items were retained. Therefore, 556 new stimulus words were needed. In finding those new words, the authors examined the initial list of 3,885 words from which the PPVT was developed and selected some possibilities from the unused portion. Webster's New Collegiate Dictionary was also scanned for possible new stimulus words, as well as word lists and studies involving surveys of words used and understood by children and youth. In assembling the pool of stimulus words, an attempt was made to get a good balance of gerunds (verbs), nouns, and descriptors. Nineteen categories were used to assure some degree of balance in selecting the new stimulus words, they are as follows: (1) actions (gerunds only); (2) animals;

(3) buildings; (4) clothing and accessories; (5) descriptors; (6) foods; (7) household and yard fixtures; (8) household utensils; (9) human body parts; (10) human workers and other role player; (11) human and humanoid forms; (12) mathematical items; (13) plants and their parts; (14) produce; (15) school and office supplies and equipment; (16) tools, machinery, scientific apparatus, and their parts; (17) toys, musical instruments, and recreational items; (18) vehicles and other means of transportation; and (19) weather, outdoor scenes and objects, and geographical items. The item search process resulted in a total of 684 stimulus words, 16 less than the goal (Dunn & Dunn, 1981).

According to the manual (Dunn & Dunn, 1981) the median coefficients of internal consistency reliability, using the split-half method, for children and youth ages 2 1/2 through 18 years, were .80 on Form L and .81 on Form M. Also, the manual reports reliability coefficients for the raw scores and standard scores. The median immediate retest alternate-forms reliability coefficient for the raw scores was .82 and for the standard scores .79; the median delayed retest alternate-forms reliability coefficients for raw scores was .78 and for standard scores .77. The delay between tests was a minimum of 9 days and a maximum of 31 days.

Tillinghast, Jr., Morrow, and Uhlig (1983) conducted a study in order to obtain additional information concerning the reliability of the PPVT-R to supplement the data presented in the manual. Tillinghast et al. randomly selected 120 pupils from grades 4, 5, and 6 and tested 40 pupils in each grade. Pearson product-moment coefficients of correlation were computed for each grade from the raw scores, and alternate-form reliability coefficients compared favorably with those reported in the PPVT-R manual. Finally, the authors combined Forms L and M into one longer test and found the retest reliability coefficients to be .91, .95, and .93 for grades 4, 5, and 6. The authors feel that there is an advantage in increasing the reliability by administering both forms as a single combined test.

Also, when investigating the alternate form equivalency of the PPVT-R for white and black preschool children, McCallum and Bracken (1981) found that both forms could be used interchangeably for most preschool children. Furthermore, Bing and Bing (1984) found that Forms L and M are equivalent for rural black preschoolers and that they can also be used interchangeably.

Beginning with the first published PPVT, various studies have been conducted to compare the scores obtained on the PPVT with the scores obtained on the CMMS. However, only two studies were found comparing the revised PPVT-R with the CMMS.

Carvajal, McVey, Sellers, Weyand, and McKnab (1987) studied the relationships between the scores on the General Purpose Abbreviated Battery of the Stanford-Binet IV, PPVT-R, CMMS (1972), and Goodenough-Harris Drawing Test (GHDT). Carvajal et al. (1987) tested a third grade class of 23 children from a midwest community of about 27,000 giving each child all four tests. The correlations between CMMS, GHDT, and PPVT-R were not statistically significant, ranging from .222 to .250. However, the correlation between the Binet IV and the PPVT-R was .601, suggesting the PPVT-R is a good screening instrument. On the other hand, the Binet IV and the CMMS had a lower correlation (.477). The results from this study suggest that the CMMS should be used sparingly, and the PPVT-R may be the better screening test.

In another study by Carvajal, Hardy, Harmon, Sellers, and Holmes (1987) the relationships among scores on the Stanford-Binet IV, PPVT-R, and CMMS (1972) were studied, and the correlations were consistent with Carvajal, Mcvey et al. (1987) previous study. From a group of 21 kindergarten children, each of which were given all three of the above tests, the correlation between the PPVT-R and the CMMS was .222. The Binet IV and CMMS had a correlation of .40, and the Binet IV and PPVT-R correlation was .56.

The CMMS and the PPVT-R have also been used for testing mentally retarded children. Ritter, Duffey, and Fischman (1974) tested 45 educable mentally retarded children ages 4 years, 11 months to 9 years, 10 months with the CMMS (1972) and Stanford-Binet, Form L-M. The correlation found between the two tests was .74 which suggests good concurrent validity even though the verbal and nonverbal requirements of the two tests are different. This correlation is consistent with the manual (Burgemeister et al., 1972). This study defends the use of the CMMS as a screening instrument.

Riviere (1973) tested institutionalized mentally retarded children with the 1972 Stanford-Binet (S-B) the CMMS (1972), PPVT, and the Wechsler Intelligence Scale for Children (WISC), and found the correlation between the S-B and the CMMS to be .52, and the S-B and the PPVT .58. The WISC correlated .437 with the CMMS and .505 with the PPVT. Finally, the PPVT and the CMMS correlated .463. This correlation was consistent with what Pascale (1973) previously reported (.47). All correlations were statistically significant.

Finally, Johnson and Shinedling (1974) administered the CMMS, PPVT, and the Slosson to 96 mentally retarded subjects ranging from 6 to 18 years. The correlations are as follows: the CMMS and the PPVT correlated .77, the CMMS and the Slosson .82, and the PPVT and the Slosson .90. Both the CMMS and the PPVT appear to be good screening tests for mentally retarded children; nevertheless, in this study and the previous one (Riviere, 1973) the PPVT correlates higher than the CMMS with various intelligence tests.

The following three studies report the use of the CMMS and the PPVT with cerebral palsied children. Nicholson (1970) investigated the characteristics of scores on the PPVT, CMMS, and the Raven's Coloured Progressive Matrices (RCPM) for cerebral palsied children. Twenty boys and 18 girls were given each test. The correlations were as follows: the correlation between the CMMS and the PPVT was .65, the CMMS and the RCPM was .74, and the PPVT and the RCPM was .41. These data indicate the scores on these three tests do correlate significantly and positively for cerebral palsied children. However, the mean IQ of the PPVT was approximately 20 points above the mean of the other two tests. The differences among the tests suggest that they measure various skills differently.

Dunn and Harley (1959) investigated the scores among the PPVT, the Ammons, the Van Alstyne, and the CMMS (1950). They found that the intercorrelations between tests generally cluster around .90 for picture vocabulary scales, while these scales correlated in the area of .80 with the CMMS. The author points out that the CMMS may be assessing intellectual factors other than those tested by the picture vocabulary tests. The results indicate that all four tests were successfully used with cerebral palsied children.

Coop, Eckel, and Stuck (1975) administered the Pictorial Test of Intelligence (PTI), the PPVT, and the CMMS, to 46 children. These children were between the ages of four and seven, and they all had cerebral palsy. The correlation between the PTI and the PPVT was .83, and the PTI and the CMMS was .88. Both correlations were significant. However, the PTI proved to be a better predictor of academic achievement than the other two tests.

Finally, the CMMS (1959) as well as the PPVT (1959) underestimated the intelligence of poverty-area first grade and kindergarten children (Rosenberg & Stroud, 1966). In the first study, 28 Negro children in a kindergarten class of a poverty-area school were tested with the PPVT, the CMMS (1959), and the Stanford-Binet Intelligence Scale, Form L-M (S-B). The data demonstrated that those labeled "Mental Defective" by the PPVT and the CMMS were children whose intelligence according to the S-B fell somewhere between 90 to 109. A second study was conducted with first grade children. It was concluded that the degree of error is reduced somewhat with children who have undergone schooling; however, the PPVT and the CMMS still greatly underestimated intelligence in poverty-area children.

After examining the previous articles, it appears that the PPVT-R does not correlate well with the CMMS (1972). However, when testing mentally retarded and cerebral palsied children, the PPVT and the earlier editions of the CMMS seem to correlate very well. The earlier editions of the CMMS also seem to correlate much better with the Stanford Binet Form L-M than it does with the Binet IV.

Purpose of the Study

Due to lack of time and money, those involved in testing are in need of a quick, cost effective, and reliable method for assessing children's abilities. Both the CMMS and the PPVT-R serve this purpose. They are very popular, frequently used, and the outcomes of these two tests can be used to determine if a more comprehensive assessment is needed. Because of this, those who use these tests will want to know how well they correlate with each other. The purpose of the current study is to determine the correlation between the CMMS and the PPVT-R.

CHAPTER 2

Method

Subjects

This study included 51 children from a midwest community of approximately 27,000. The target population consisted of children attending Butcher Elementary School, which is associated with Emporia State University, in Emporia, Kansas. Any student within Unified School District #253 is eligible to attend Butcher Elementary School. The sample included students ranging in age from 3 years to 7 years, who were enrolled in preschool, kindergarten, first, or second grade.

Letters were sent to all the parents of the subjects, and they were asked to read and sign an informed consent form before the study was conducted. The form explained both the purpose and the procedures of the study. Those children whose parents signed the form verified agreement for the child to be tested.

Confidentiality was ensured accordingly. An application was submitted to Emporia State University's Review Board for Treatment of Human Subjects prior to the mailing of letters to the parents.

Procedure

The Peabody Picture Vocabulary Test-Revised (PPVT-R) and the Columbia Mental Maturity Scale (CMMS) were individually administered to each subject. The tests were administered by two faculty members from the Division of Psychology and Special Education at Emporia State University. This prevented experimenter bias. The tests were administered in a counterbalanced sequence, and all tests were given under standard conditions and followed standardized procedures.

<u>Statistical Design</u>

The type of research in this study described was correlational research. This study examined the relationship of the two variables, the CMMS and PPVT-R. This was done by computing the Pearson correlation coefficient between the scores yielded by each test.

This type of research is described as a concurrent criterion-related validity study. Bailey and Wolery (1989) state that "concurrent validity refers to the extent to which a test correlates with another measure administered close in time to the first" (p. 42).

This study produced two scores per subject: one from the PPVT-R referred to as a Standard Score Equivalent (SSE) and one from the CMMS referred to as and Age Deviation Score (ADS). To determine the association between the PPVT-R scores and the CMMS scores, they were analyzed by calculating the Pearson product-moment correlation coefficient (Pearson $\underline{r}$). Descriptive statistics for the entire sample and gender groups for both tests were reported and analyzed by comparing them with the normative data.

Also a 2 X 2 mixed factorial analysis of variance (ANOVA) was performed to determine the difference between gender on the test scores, the difference between the scores on the PPVT-R and the CMMS, and the interaction between the two tests and gender. The between groups are girls and boys and the within groups are the PPVT-R and CMMS. Because the standard deviations on the CMMS and PPVT-R were not the same (i.e., 16 and 15, respectively) the mean scores were converted to z scores before performing the ANOVA.

CHAPTER 3

RESULTS

One score from the Columbia Mental Maturity Scale (CMMS) and one from the Peabody Picture Vocabulary Test-Revised (PPVT-R) were obtained for 51 children (27 boys and 24 girls). The children ranged in ages from 3 years, 6 months to 7 years, 3 months. The means, standard deviations, ranges, minimum scores, and maximum scores are presented in Table 1.

Table 1

Descriptive Statistics for the CMMS and PPVT-R

Test	*M*	*SD*	Range
CMMS			
ADS	110.52	9.36	45 (96-141)
PPVT-R			
SSE	110.98	12.11	57 (79-136)

The mean SSE score reported by this study was 110.98 (*SD* = 12.11) for the PPVT-R. This mean score was higher than the normative groups for the PPVT-R which reported a mean of 100 (*SD* = 15). The mean CMMS ADS score was 110.52 (*SD* = 9.36) which is also higher than the normative group which reported a mean of 100 (*SD* = 16). Furthermore, after observing the standard deviation scores, the variability of the sample is smaller than the normative group. This sample is more homogenous than the normative sample. Tables 2 and 3 list the descriptive statistics of the subjects by gender.

Table 2

Descriptive Statistics for the CMMS and PPVT-R for Females

Test	*M*	*SD*	Range
CMMS			
ADS	112.50	11.88	45 (96-141)
PPVT-R			
SSE	107.95	11.71	54 (79-133)

Table 3

Descriptive Statistics for the CMMS and PPVT-R for Males

Test	*M*	*SD*	Range
CMMS			
ADS	108.77	6.35	21 (97-118)
PPVT-R			
SSE	113.66	12.26	48 (88-136)

The mean scores reported in Tables 2 and 3 are above the normative samples for both the CMMS and PPVT-R. The females scored higher than the males in regards to mean scores on the CMMS. The females mean score was 112.50 (*SD* = 11.88) and the males 108.77 (*SD* = 6.35). However, the males scored higher than the females in regards to mean scores on the PPVT-R; the males scoring 113.66 (*SD* = 12.26) and the females 107.95 (*SD* = 11.71).

A Pearson product-moment correlation coefficient was calculated between the CMMS and the PPVT-R. The CMMS and the PPVT-R correlated very poorly ($r = .18$).

A 2 x 2 factorial analysis of variance (ANOVA) was conducted to determine if there were gender differences between the CMMS and the PPVT-R. No main effects were obtained for Gender or Tests. However, the Gender X Tests was significant, $F(1,49) = 6.26$, $p < .05$. Fisher's Least Significant Difference Test was applied to determine where the difference was. The males score on the PPVT-R was significantly higher than the females score on that test. Also, the males score on the PPVT-R was significantly higher than the males score on the CMMS ($p < .05$).

CHAPTER 4

DISCUSSION

It is very important for individuals involved in testing to have the most valid instrument available. Important academic decisions are made on the basis of such test scores. For instance, screening tests are vitally important in order to determine whether or not one may need a further and more complete assessment. If a screening instrument is not able to identify those individuals, they may be placed in an improper academic setting. Screening instruments can also help school personnel make decisions concerning the selection of curriculum materials and learning tasks that have been made available for all the children. When using a screening instrument for assessing children it is necessary to be aware of what abilities the instrument is assessing. One needs to be very careful not to overgeneralize the score from a single-skilled test because usually only one skill is tested, as opposed to using a multi-skilled instrument. The present study was designed to determine how the Columbia Mental Maturity Scale (CMMS) and the Peabody Picture Vocabulary Test-Revised (PPVT-R), two screening instruments, correlate.

In the present study, overall the two instruments produced almost identical mean ADS/SSE scores. However, the males scored much better on the PPVT-R than the females, and the females scored better on the CMMS than the males. This finding is somewhat surprising. One would think males and females would score about the same; however, the differences found in this study may be attributed to the sample. The sample may not be representative of the entire population since the children were selected from one institution. There was also a significant difference by gender. The males score on the PPVT-R was significantly higher than the females score on the PPVT-R, and the males score on the PPVT-R was significantly higher than the males score on the CMMS. This difference may also be due to the sample used.

Even though the overall mean scores on the two tests were very similar, the correlation coefficient between the scores of the two tests was not statistically significant. The similar mean scores suggest that the two tests are capable of producing comparable assessments. Nevertheless, the low correlation coefficient demonstrates that the two tests do not produce comparisons of the same equality. One should not substitute the results of one test for the other. This low correlation is not surprising and probably occured because each test assesses a different intellectual task. For instance, the PPVT-R measures receptive hearing vocabulary and the CMMS measures general reasoning ability. This low correlation is consistent with previous research involving sample sizes from the same institution.

Both the CMMS and the PPVT-R are screening instruments and should not be used as a replacement for each other because they measure different intellectual abilities. The individuals involved in testing need to be aware of what skills a particular test is assessing, and they need to know whether or not one particular test can be substituted for another. For the CMMS and the PPVT-R to accurately assess individuals, they should be used to assess only the abilities they were designed to test. Those abilities are general reasoning ability (CMMS) and receptive hearing vocabulary (PPVT-R). Furthermore, the ultimate value of a screening device is determined by its relationship with a major test, such as the Stanford-Binet or the Wechsler series. Both the CMMS and the PPVT-R have been compared to such instruments.

Further research needs to be conducted using both "regular" and handicapped children. Past literature demonstrates that the CMMS and the original PPVT (1959) produced a statistically significant correlation coefficient when testing handicapped children. It would be interesting and beneficial to demonstrate how the CMMS and the revised PPVT (1981) correlate using handicapped children.

Two very important issues that should be considered for future research are as follows. Samples should be drawn from a more diverse population, and more individuals should be included in the sample size.

REFERENCES

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

Bailey, D. B., Wolery, M. (1989). Assessing infants and preschoolers with handicaps. Columbus, OH: Merrill Publishing Company.

Bing, J. R., & Bing, S. B. (1984). Alternate-form reliability of the PPVT-R for black head start preschoolers. Psychological Reports, 54, 235-238.

Burgemeister, B. B., Blum, L. H., & Lorge, I. (1972). Guide for administering and interpreting the Columbia Mental Maturity Scale (3rd ed.). New York: Harcourt Brace Jovanovich.

Carvajal, H., Hardy, K., Harmon, K., Sellers, T., & Holmes, C. (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.

Coop, R.H., Eckel, E., Stuck, G.B. (1975). An assessment of the pictorial test of intelligence for use with young cerebral-palsied children. Developmental Medicine and Child Neurology, 17, 287-292.

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

Dunn, L. M., & Harley, R.K. (1959). Comparability of Peabody, Ammons, Van Alstyne, and Columbia test scores with cerebral palsied children. Exceptional Children, 26, 70-74.

Egeland, B. R. (1978). Review of Columbia Mental Maturity Scale. In O. K. Buros (Ed.), The eighth mental measurements yearbook (pp.298-299). Highland Park, NJ: The Gryphon Press.

Hiskey, M. S. (1965). Review of Columbia Mental Maturity Scale. In O. K. Buros (Ed.), The sixth mental measurements yearbook (pp. 800-803) Highland Park, NJ: The Gryphon Press.

Johnson, R. L., & Shinedling, M. M. (1974). Comparison of Columbia Mental Maturity Scale, Peabody Picture Vocabulary Test, and Slosson Intelligence Test with mentally retarded children. Psychological Reports, 34, 367-370.

Kaufman, A. S. (1978). Review of Columbia Mental Maturity Scale. In O. K. Buros (Ed.), The eight mental measurements yearbook (pp. 299-301). Highland Park, NJ: The Gryphon Press.

Lubin, B., Larsen, R. M., Matarazzo, J. D., & Seever, M. (1985). Psychological test usage patterns in five professional settings. American Psychologist, 40, 857-861.

McCallum, R. S. (1985). Review of Peabody Picture Vocabulary Test- Revised. In J. M. Mitchell (Ed.), The ninth mental measurements yearbook (pp. 1126-1127). Lincoln, NE: Buros Institute of Mental Measurements of the University of Nebraska-Lincoln.

McCallum, R. S. & Bracken B. A. (1981). Alternate form reliability of the PPVT-R for white and black preschool children. Psychology in the schools, 18, 422-425.

Nicholson, C. L. (1970). Correlations among CMMS, PPVT, and RCPM for cerebral palsied children. Perceptual and Motor Skills, 30, 715-718.

Pascale, P. J. (1973). Validity concerns of preschool testing. Educational and Psychological Measurements, 33, 977-978.

Roberson, G. J., & Eisenberg, J. L. (1981). Technical Supplement. Circle pines, MN: American Guidance Service.

Ritter, D., Duffey, J., & Fischman, R. (1974). Comparability of Columbia Mental Maturity Scale and Stanford-Binet, Form L-M estimates of intelligence. Psychological Reports, 34, 174.

Riviere, M. S. (1973). The use of the Columbia Mental Maturity Scale with institutionalized mentally retarded children. Educational and Psychological Measurement, 33, 993-995.

Rosenberg, L. A., & Stroud, M. (1966). Limitations of brief intelligence testing with young children. Psychological Reports, 19, 721-722.

Tillinghast, B. S., Morrow, J. E., & Uhlig, G. E. (1983). Retest and alternate-form reliabilities of the PPVT-R with fourth, fifth, and sixth grade pupils. Journal of Educational Research, 76, 243-244.

Wiig, E. H. (1985). Review of Peabody Picture Vocabulary Test-Revised. In J. M. Mitchell (Ed.), The ninth mental measurements yearbook (pp.1127-1128). Lincoln, NE: Buros Institute of Mental Measurements of the University of Nebraska-Lincoln.