A COMPARISON OF THE CALIFORNIA TEST OF MENTAL MATURITY AND THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN IN FOUR CLINICAL GROUPS OF SCHOOL CHILDREN

APPROVED:

Earl W Kooker
Major Professor

Jack R. Hayes
Minor Professor

Duane Kingery
Dean of the School of Education

Robert B. Toulouse
Dean of the Graduate School

A COMPARISON OF THE CALIFORNIA TEST OF MENTAL MATURITY AND THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN IN FOUR CLINICAL GROUPS OF SCHOOL CHILDREN

THESIS

Presented to the Graduate Council of the
North Texas State University in Partial
Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

by

Leslie A. Nichols, B. S.

Denton, Texas

August, 1969

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

The use of group intelligence tests in school settings is an accepted practice. Likewise, it is not uncommon for school personnel to refer a child to an educational clinic for individual testing when he presents learning or behavior problems. It has been observed, however, that large discrepancies often appear between the results of the group and individual intelligence tests for children referred for individual evaluation, and it is this phenomenon with which this study was concerned.

Two tests which are employed widely in the above contexts are the California Short-Form Test of Mental Maturity and the Wechsler Intelligence Scale for Children. A brief discussion of each of these tests is presented below.

The California Test of Mental Maturity (CTMM) was developed by Sullivan, Clark, and Tiegs, and was originally published in 1936. The test was developed in response to the need for a less time consuming and expensive method of assessing intelligence in school settings than was possible

with the Stanford-Binet (Binet). It was desired, however, to construct a scale that would parallel the individual Binet as closely as possible, and the authors applied the conceptual framework and rationale of the Binet in preparing items. Both scales utilized the mental age concept, had standard deviations of 16 I. Q. points, and purportedly tested the same type of mental abilities.

The original CTMM consisted of four logical constructs which were called factors. They were designated Memory, Spatial Relationships, Reasoning, and Vocabulary. There were also three pretests, Visual Acuity, Auditory Acuity, and Motor Coordination, and the test spanned five school levels: Pre-Primary, Primary, Elementary, Intermediate, and Advanced.

A number of changes have taken place in the CTMM since the original edition, and the most important of these were as follows: (1) the publication of the first California Short-Form Test of Mental Maturity in 1938 in response to the demands of school personnel for a one-period test that would embody the major features of the regular CTMM; (2) removal of the pre-tests from the body of the test in 1950; (3) restandardization, the modification of existing levels, and the addition of a new one for high school in 1957;

(4) in 1963, restandardization, revision of test content, the addition of two more levels to the Short-Form, scaling to the Stanford-Binet Intelligence Scale, Form L-M, 1960 Revision, and the replacement of the ratio IQ with the deviation IQ.

The 1963 edition of the CTMM series is the most recent, and as the California Short-Form Test of Mental Maturity (CTMM S-F), 1963 Revision, is the edition with which this study was concerned, the following description will be limited to this particular series.

In their present form, these series offer eight articulated levels ranging from preschool to adult. These various levels are designated Level 0 through Level 5, and there is some overlap among them with respect to age and grade coverage. The purpose of such overlap is to provide flexibility for differing school needs, and it is not necessary to give all levels in order to span the range of ages.

Each test yields a Language IQ (LIQ), Non-Language (NLIQ), and Total IQ, and the range of intellectual functioning to which normative data are most applicable extends from IQs of 63 to 137. Special tables are included in the manuals for functioning below or above these limits,

but the user is cautioned that they may not be as reliable as the regular tables. In addition to IQs, the raw scores are also expressible "in percentile ranks, two kinds of standard scores, standard score IQs, MAs, and an ISI (Intellectual Status Index, which reflects the examinee's performance in relation to a national norm population for his grade placement)" (19, p. 694).

Total testing time is approximately 40 minutes, and there are seven timed subtests for each Level. These subtests are: (1) Opposites; (2) Similarities; (3) Analogies; (4) Numerical Values; (5) Number Problems; (6) Verbal Comprehension; and (7) Delayed Recall.

The factors represented by these subtests, while similar to those in the original CTMM, have certain refinements: they were derived by the Thurstone centroid method and are now called Factor I, Logical Reasoning; Factor II, Numerical Reasoning; Factor III, Verbal Concepts; and Factor IV, Memory. Subtests one through three comprise Factor I, subtests four and five Factor II, subtest six Factor III, and subtest seven Factor IV. The Language IQ is composed of subtests five, six, and seven; the Non-Language IQ of subtests one, two, three, and four.

The CTMM is generally administered to students by a teacher, and "directions for administering the various levels are for the most part clear, simple, and detailed" (19, p. 694). Examinees are required to mark their answers on standardized forms, and reading ability is not required at the lowest levels. Beginning at Level 2, however, there is progressively more material requiring ability to read. The 1963 Short-Form may also be administered via tape recorder.

As reported in the Technical Report (TR) (7) published by the California Test Bureau, reliability coefficients for levels relevant to this study ranged from .90 to .95 for Language IQs, from .81 to .89 for Non-Language IQs, and from .91 to .95 for Total IQs.

Kuder-Richardson Formula 21 was used in computing coefficients for all levels except Level 1, which was computed by the split-halves method. Reliability coefficients for Level 0 were somewhat lower (.70, 159, 178), but this study was concerned only with Levels 1 through 3. The standard errors of measurement for levels used in this study were reported to range from 3.8 to 5.2 for Total IQs.

Correlations with the criterion instrument, the Stanford-Binet Intelligence Scale, Form L-M, 1960 Revision,

(corrected for range and attenuation), were reported in the TR to be in the .80s for all levels of the CTMM relevant to this study. Correlations with other tests (Otis, Henmon Nelson, SCAT) were also presented in the TR, and it was noted that none fell below .78, while coefficients indicating concurrent validity with the 1957 Short-Form were reported to range from .70 to .91 for levels pertinent to this study. No independent indices of predictive validity were offered in either the manual or the TR for the 1963 Revision, although high correlations were reported between the California Achievement Test and the CTMM S-F, 1963 Revision. As this achievement test was developed in conjunction with the intelligence test, high correlations might obtain, irrespective of their actual predictive validity as measured by more independent criteria.

Concerning the standardization of the 1963 CTMM series, the following has been reported:

> The sample utilized in the scaling and norming of the California Test of Mental Maturity series, 1963 Revision, was drawn in such a manner as to establish norms applicable to the national population. Cases were obtained by testing independent class samples from 253 schools selected from seven geographic regions, representing 49 states. Participating schools were selected with regard to the factors of geographical region, school population, school grade, and community size. General educational, social, cultural, and

> economic conditions were considered in grouping states with similar characteristics into seven regions.
>
> In the norming of instruments such as the Short-Form which is scaled directly to another instrument, the emphasis is upon the scaling procedure and not upon selecting a truly random or rigidly controlled, representative sample. The sample was drawn from the seven initially established areas in a manner so as to be roughly proportional to the total school population, but the aspect of scaling the Short-Form Total Raw Score to the Stanford-Binet IQ assumed the greater emphasis in developing the 1963 Revision (7, p. 10).

With respect to the frequency with which the CTMM series are utilized in school settings, there is evidence that it is one of the most popular and widely used group tests. Cronbach stated it is

> One of the most widely accepted current tests, with unusual variety of items, good format and standardization, and a continuous series of levels. There is a Short-Form for use where less reliable measurement is acceptable (9, p. 229).

In concluding the discussion and description of the CTMM per se, Altus offered this interesting observation:

> One of the reasons for the increasing preference for the WISC over the Stanford-Binet is undoubtedly its separation into Verbal and Performance scales. Similarly, choice of the California Test of Mental Maturity, or CTMM, as a group measure of intelligence is frequently made because it yields both Language and Non-Language IQs. The tacit assumption is often made, as by school psychologists using both tests, that the corresponding verbal and non-verbal portions of the two tests are highly correlated (2, p. 143).

Wechsler Intelligence Scale for Children

The Wechsler Intelligence Scale for Children (WISC) was developed by Wechsler and first published in 1949. It grew logically out of the Wechsler-Bellevue Intelligence Scales which Wechsler had constructed in 1939 in response to a need for an individual test of intelligence for patients at Bellevue Hospital in New York. The Wechsler-Bellevue scales are now obsolete and have been replaced by better constructed and better standardized forms. These forms are the above-mentioned WISC, the Wechsler Adult Intelligence Scale (1955), and the Wechsler Preschool and Primary Scale of Intelligence (1967).

In the development of his tests, Wechsler subscribed to the Binet idea of general mental ability, but "sought items, which, while falling within the area we identify as general mental ability, had sufficiently specific characteristics to silhouette different types of thinking and performance" (9, p. 192). Further, Wechsler defined intelligence as "the global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment" (22, p. 7).

The WISC is intended to reflect this rationale, and is composed of ten different major subtests: (1) Information;

(2) Comprehension; (3) Arithmetic; (4) Similarities; (5) Vocabulary; (6) Picture Completion; (7) Picture Arrangement; (8) Block Design; (9) Object Assembly; and (10) Coding. The first five subtests constitute the Verbal Scale, and the latter five the Performance Scale. Additionally, there is an optional subtest for the Verbal Scale (Digit Span), and an optional subtest for the Performance Scale (Mazes).

The WISC yields a Verbal Scale IQ (VSIQ), a Performance Scale IQ (PFIQ), and a Full Scale IQ (FSIQ). Total time for administration of the WISC is approximately one hour, and the test is intended to be administered only by psychologists or other appropriately qualified and trained persons.

Wechsler (21, p. 113) utilized the deviation IQ, eschewing the mental age concept, although he did provide a table of "Test Age Equivalents for WISC Raw Scores" which approximates the mental age concept. The mean and standard deviation of the WISC are 100 and 15, respectively. Age levels appropriate to the WISC are five through fifteen, and the standardization of the instrument included IQs ranging from 46 to 154.

Reliability coefficients, as reported in the manual, range from .88 to .96 for the Verbal Scale, .86 to .90 for

the Performance Scale, and from .92 to .95 for the Full Scale. Wechsler (21, p. 13) reported the split-half method was used in conjunction with the Spearman-Brown formula for correction for length. No concurrent or predictive validity data were given in the manual.

The standardization sample consisted of 2200 white children, 1100 boys and 1100 girls in eleven age groups (100 in each age group). The standard errors of measurement for three age groups (7½, 10½, and 13½) were, respectively, reported as 4.25, 3.36, and 3.68 for Full Scale IQs.

In concluding the description of the WISC this comment by Burstein seems appropriate:

> In an era when fads in test construction and test consumption combine to produce rapid obsolence and turnover, the WISC can be regarded as a highly successful test, if only on the grounds of durability. In the nearly 15 years since its introduction, it has not displaced the older Stanford-Binet, but has certainly come to rival its predecessor as an instrument of choice in the testing of school age children (6, p. 843).

Related Research

Data reported in the TR bearing on the reliability and concurrent validity of the CTMM series with which this study was concerned have already been cited in the section describing the CTMM. In the interest of objectivity, however,

several more studies will be mentioned which pertain to the validity and reliability of the CTMM series as it has sometimes been observed that test publishers are occasionally wont to report those studies whose results reflect more favorably on their instrument.

Tatham and Dole, in a study involving two samples of above average pupils at the University of Hawaii Elementary School who ranged in age from eight to twelve, concluded the the CTMM S-F, 1957 Revision predicted IQ on the Binet, Form L-M with only moderate efficiency, and that "for individual clinical referrals CTMM S-F IQ can not be considered interchangeable with the Binet" (20, p. 302). Correlational coefficients of .41 and .56 were found between the two instruments in this study; attenuated range was postulated as a factor possibly lowering coefficients.

Rainey (16) designed a study to evaluate four school ability tests. the Lorge-Thorndike, 1957 Edition; the Kuhlmann-Anderson, 1963 Edition; the California Short-Form Test of Mental Maturity, 1963 Edition; and the Primary Mental Abilities Test, 1963 Edition. Three groups composed of second, fourth, and seventh graders who were further subdivided into three socio-economic levels for each grade were given the tests, and with respect to the CTMM, the author

concluded that the mean and standard deviation stability of the CTMM was good only in the high socio-economic area, producing the highest means in this area of the four tests and the lowest of the four tests in the low socio-economic area, as well as the greatest variability of the four tests in this latter area.

Finley, Thompson, and Cognata (11) attempted to measure the stability of the CTMM S-F, 1957 Revision for grades three, five, and seven in a longitudinal study involving some 314 pupils in the Sonoma County California Schools, and found significant (.05 level) drops in Language IQs between grades three and five. Scores tended to rise again in grade seven, however. Perhaps the most significant finding of this study was that "the amount of individual variation which can be expected (due to test unreliability) on re-test is considerable. An actual difference of 20 or more IQ points is needed before one can realistically begin to hypothesize true change or question test administration procedures" (11, p. 160).

Bradshaw (4) in a similar study found significant changes (.05 level) occurring between second and fourth grade Non-Language IQs, with 31 per cent of the population of his study obtaining Non-Language IQ differences of ± 24

points on re-test with the CTMM S-F, 1957 Revision. Sizeable differences were also found between Total IQs upon re-test (± 16 points for 28 per cent of the study population), although they were not significant at the .05 level. Correlations between subsequent testings were .65 for Language IQs, .50 for Non-Language IQs, and .66 for Total IQs.

Sheldon and Manolakes (18), in a study involving 422 pupils in grades one through six in eight schools, administered the Stanford-Binet, Form L, and the CTMM S-F within a four month period. They found that although there were no significant differences between the means of the two tests, "13.7 per cent of the cases showed differences greater than 20 points" (18, p. 500).

Thus it would appear that there is some question as to whether the CTMM S-F is as reliable an instrument as statistical evidence presented in the TR might lead the unwary user to believe. Further, a major criticism is that there were no data presented in the TR bearing directly on IQ stability of the CTMM series.

The next section of the review of the literature will present data treating of the relative merits of the WISC, especially as is related to reliability and concurrent validity. In view of the role the Stanford-Binet played in

the development of the CHC series, there would seem to be some rationale for citing some studies concerned with the WISC and the Binet, and these studies are presented initially.

Harlow, Price, Tatham, and Davidson (14), in a study involving three age groups of 90 white school children between the ages of six years six months and fourteen years three months, found correlations significant at the .01 level for all age groups among VSIQs, PSIQs, and FSIQs and the Binet, Form L. They also found that the brighter children at all age levels tested higher on the Binet than on the WISC, that the duller children tended to test slightly lower on the WISC, and that the two tests are most comparable in the 91-100 IQ range for all ages included in their study.

Freeman (8, p. 273) reported correlations between the WISC and the Binet, Form L for mentally defectives and normals that ranged from .75 to .90 for FSIQs, from .65 to .90 for VSIQs, and from .50 to .75 for PSIQs. Further, it was noted that "the Wechsler scale tends to rate subnormal subjects somewhat higher, but not markedly so, than does the Stanford-Binet. At the average level, the reverse is true" (8, p. 273). It will be remembered that the findings of Harlow, et al. differed from these.

Arnold, [illegible] a study comparing the Binet, Form L and the WISC for a random sample of 50 eight and nine year old school children, found high correlations between the two instruments (VSIQ .88, PSIQ .74, FSIQ .90), but using the correlation between Form L and Form M of the Binet (r = .93) as criterion, concluded:

> So far as this sample is concerned, the relationship between IQs obtained for eight and nine year olds with the WISC and the Form L Binet is not significantly different from the relationship between IQs obtained on Form L and M of the Binet. So far as total score is concerned then, the WISC might very well be substituted for the Binet or the Binet for the WISC. From results of this study, the same would seem to be true for the WISC Verbal Scale. This would not seem to be true, however, for the WISC Performance Scale since the relationship found differs significantly from that between Form L and M at the 1 per cent level of confidence (3, p. 93).

Schachter and Apgar (17) compared preschool Stanford-Binets, Form L with school age WISCs and obtained correlational coefficients of .64, .48, and .67, all significant at the .01 level for WISC VSIQs, PSIQs, and FSIQs and the Binet respectively. The highest correlations obtained between the FSIQs and the Binet, with VSIQ and PSIQ correlations consecutively lower, and mean Binet IQs approximately five points higher than those of the WISC. An analysis of [illegible]ariance indicated no intertest differences due to age or sex. The authors summarized several studies

relating the two instruments and concluded that their findings were in accord with the results of other studies. The following points are quoted from their summary of the other studies:

> The median correlation between the Stanford-Binet and the WISC Full Scale was .85. Highest intertest correlations obtained for the WISC Full Scale, next highest for the WISC Performance Scale. Mean Stanford-Binet IQs were significantly higher than mean WISC IQs. Significantly greater intertest discrepancies occurred at the high IQ and low age levels (17, p. 320).

Littell, in The Wechsler Intelligence Scale for Children: Review of a decade of research, stated the following conclusions regarding the WISC and the Binet, Form L:

> 1. Studies involving a variety of ages and IQ ranges are very consistent in showing that at least within a white American school population on the WISC and the Stanford-Binet scores are related to a significant degree. Correlations between the WISC Full Scale and the S-B are predominantly reported within the 80s.
>
> 2. The WISC scores tend to be lower than the S-B scores for the same children within the middle and upper ranges and somewhat higher for defectives. This appears to be particularly true for younger children (below 10) and for the higher S-B scores.
>
> 3. Using the S-B as a criterion, the highest correlations are found with the Full Scale IQ scores, the next highest with Verbal, and the lowest with Performance scores (15, p. 138).

Other noteworthy points stated by Littell were these:

> 1. There is strong evidence that WISC norms are not applicable to children of markedly different

subgroups such as southern Negro and bilingual Mexican-American children.

2. Socio-economic status appears to be a significant variable affecting the IQ scores of young children (second as opposed to fifth grade children), such that the children of higher socio-economic status tend to obtain higher scores.

3. When the WISC is administered to children with "mental ages" below five or six years, the IQ scores can be expected to be relatively unreliable due to limited number of "functional" test items at the low end of the scale (15, p. 148).

4. While the studies reviewed all report rather high correlations between the WISC and group intelligence tests, again the small number of studies precludes more than the very tentative acceptance of these conclusions (15, p. 141).

As may have been observed, no studies have yet been reported bearing on the comparability of the WISC and the S-B, Form L-M, 1960 Revision. A thorough search of the literature disclosed only one such study.

Estes, Curtin, DeBurger, and Denny compared the S-B, Form L-M, 1960 Revision, the S-B, Form L, 1937 Revision, the WISC, Raven, and Draw-A-Man, and in summary stated:

1. The comparability of IQs from five different intelligence tests were investigated for an above average group of white American children.

2. For the entire group (N = 82), scores for the 1960 S-B, the 1937 S-B, and the WISC were found to be comparable and to compare favorably with representative similar findings.

3. The age factor, contrary to some previous findings, was not found to account for test discrepancies among the two S-B and WISC instruments.

> 4. Intelligence level, in agreement with previous findings, was a factor producing highly significant discrepancies at superior levels among the two S-B and WISC instruments. More investigation is needed regarding the effect of intelligence level on test comparability.
>
> 5. Correlations relating the Raven and D-A-M to the two S-Bs and WISC were significant but relatively small (10, p. 390).

Up to this point, studies have been presented bearing on the comparability of the CTMM and Binet scales, and the WISC and Binet scales. The majority of the evidence would seem to indicate that the WISC and Binet scales generally correlate highly with each other overall, and summary studies have been presented showing the general concensus of agreement and disagreement between the two instruments. Similarly, studies were reported bearing on the comparability of the CTMM and Binet scales, and although the two instruments seem to correlate well, there appears to be some question that the reliability of the CTMM is as good as data in the TR indicate.

In searching for studies comparing the CTMM series and the WISC, a situation similar to that encountered while searching for data bearing on the 1960 Revision of the Binet and the WISC was found. Only two studies comparing the CTMM and the WISC was discovered, and neither of these concerned the 1963 Revision of the CTMM S-F.

Altus (1, in 1952, found a correlation of .81 between the CTMM S-F and the WISC for total IQs for a sample of 55 Junior High School children, and in 1955, conducted a study relating the CTMM S-F and the WISC for a population of 100 elementary school children referred to the Guidance Department of the Santa Barbara County Schools by their teachers. The CTMM S-F was given routinely by teachers as part of the county testing program, while the WISC was administered by the author. Sixty-six of the subjects were given the two tests within a period of one year, and in no case were the two tests given as much as three years apart. Correlations obtained ranged from .57 for the Language CTMM and the Performance WISC, to .77 for the full scales. One child in four showed a difference in excess of 18 IQ points between WISC Performance and CTMM Non-Language IQs, while the difference between total IQs on the two tests was less than 10 points in 65 per cent of the cases. Altus concluded:

> In short, it would seem justified to conclude that, within a comparable school referral setting, the WISC and CTMM are markedly comparable as to group assessment and roughly comparable as to individual scores and major breakdown into verbal and nonverbal abilities (2, p. 144).

It will be recalled that Littell pointed out that there may be dangers involved in accepting, as more than tentative,

conclusions which are based on a small number of studies. Those comments made by Fraser would also seem to bear upon the dangers involved in assuming interchangeability of IQ scores between different tests.

> The putting together of the items with the greatest internal validity is well worthwhile, but there is no guarantee that the final test, taken as a whole, will give the same results as either the old Form L or Form M. Thus, even if items a + b + c + d + e + f from Form L were equivalent to items m + n + o + p + q + r from Form M, it is not necessarily true that a + b + m + d + q + r will be equivalent to either of the former groupings. This has yet to be shown to be the case (12, pp. 830-831).

The inference, as related to this study, is that if there is reason to question the comparability of seemingly highly similar tests such as the Binet, Form L and Form L-M, one should be even more cautious in assuming interchangeability of IQs of two apparently dissimilar tests such as the WISC and the CTMM. Further, the latter part of Fraser's statement would seem to indicate the need for numerous studies demonstrating the relationships between different tests before one is justified in stating what these relationships are.

Significance of the Study

As has been mentioned earlier in this study, there seems to be a general tendency for psychologists and

educators to "tacitly assume" different IQ tests are reasonably interchangeable with regard to obtained scores. When large IQ differences are obtained in subsequent testings for a given individual, either on the same or on different instruments, any or all of the following have been observed to occur: (1) there appears to be an inclination to conclude some change has taken place in the individual rather than that the test(s) may be unreliable or non-equivalent; (2) there would also seem to be a tendency to go to the other extreme and discount the efficacy of IQ tests entirely when such differences are obtained; and (3) school personnel often express bewilderment when large IQ discrepancies appear for a given individual, and understandably, are at something of a loss to decide which score should be used as a basis for academic planning.

Factors which seemingly contribute to this state of affairs are the tendencies of some test makers to publish data which reflect favorably on their instruments, the publication of insufficient data, and the lack of familiarity with the interpretation of statistical data by test users. These factors would seem especially important as pertains to the limitations of published data when applied to populations other than the normative one.

Thus if one accepts any or all of the above as valid observations, there would appear to be some merit in conducting studies which attempt to delineate some of the limitations of IQ tests more clearly. Further, it would seem especially important for educators to be familiar with such studies because of the role IQ scores play in making decisions concerning students. If, indeed, it is demonstrated that relatively large IQ differences may be expected to obtain for certain tests with specific populations, then it seems likely that IQ scores might be interpreted more realistically, and an important educative function may have been served, both to the decision makers and those whose lives are affected by these decisions.

Statement of the Problem

The primary problem of this study was to compare the WISC and CTMM S-F, 1962 Revision, in order to determine whether the two instruments were interchangeable with respect to intelligence quotients for a school-clinical population. In order to further identify the differences that might possibly exist between the two instruments for populations atypical of the normative ones, four separate groups of school children were studied.

Hypotheses

Three major hypotheses were advanced, with subjective impressions accruing from testing experiences in an educational clinic largely the basis for the second and third hypotheses:

I. Significant positive correlations would be found between the CTMM Language and WISC Verbal IQs, between the CTMM Non-Language and WISC Performance IQs, and between the CTMM Total and WISC Full Scale IQs for each of the four groups of this study.

II. There would be a significant variation among the correlations in the four groups for the variables CTMM Language and WISC Verbal IQs, CTMM Non-Language and WISC Performance IQs, and CTMM Total and WISC Full Scale IQs.

III. There would be significant mean differences in IQ scores between the CTMM Language and WISC Verbal IQs, CTMM Non-Language and WISC Performance IQs, and CTMM Total and WISC Full Scale IQs for each group.

CHAPTER BIBLIOGRAPHY

1. Altus, Grace T., "A Note on the Validity of the Wechsler Intelligence Scale for Children," Journal of Consulting Psychology, XVI (June, 1952), 231.

2. _____, "Relationships Between Verbal and Non-Verbal Parts of the CTMM and WISC," Journal of Consulting Psychology, XIX (April, 1955), 143-144.

3. Arnold, Frank C. and Winifred K. Wagner, "A Comparison of Wechsler Children's Scale and Stanford-Binet Scores for Eight- and Nine-Year Olds," Journal of Experimental Education, XXIV (September, 1955), 91-94.

4. Bradshaw, Donald H., "Stability of California Test of Mental Maturity IQ's from the Second to the Fourth Grade," Educational and Psychological Measurement, XXIV, No. 4 (Winter, 1964), 935-939.

5. Buros, Oscar K., The Sixth Mental Measurements Yearbook, Highland Park, New Jersey, The Gryphon Press, 1965.

6. Burstein, Alvin G., "Review of the WISC," The Sixth Mental Measurements Yearbook, edited by O. K. Buros, Highland Park, New Jersey, The Gryphon Press, 1965.

7. California Test Bureau Personnel, Technical Report on the California Test of Mental Maturity Series, 1963 Revision, Monterey, California, California Test Bureau, 1965.

8. Clark, Willis W., et al., California Short-Form Test of Mental Maturity: Manual (Level 2, 1963 Revision), Monterey, California Test Bureau, 1963.

9. Cronbach, Lee J., Essentials of Psychological Testing, New York, Evanston, and London, Harper and Row, 1960.

10. Estes, Betsy W. and others, "Relationships Between 1960 Stanford-Binet, 1937 Stanford-Binet, WISC, Raven and Draw-A-Man," Journal of Consulting Psychology, XXV (October, 1961), 388-391.

11. Finley, Carmen J., Jack M. Thompson, and Albert Cognata, "Stability of the California Short-Form Test of Mental Maturity: Grades 3, 5, and 7," California Journal of Educational Research, XVII, No. 4 (September, 1966), 157-168.

12. Fraser, Elizabeth D., "Review of the WISC," The Sixth Mental Measurements Yearbook, edited by O. K. Buros, Highland Park, New Jersey, The Gryphon Press, 1965.

13. Freeman, Frank S., Theory and Practice of Psychological Testing, New York, Holt, Rinehart and Winston, 1963.

14. Harlow, Justin E., Jr. and others, "Preliminary Study of Comparisons Between Wechsler Intelligence Scale for Children and Form L of Revised Stanford-Binet Scale at Three Age Levels," Journal of Clinical Psychology, XIII (January, 1957), 72-73.

15. Littell, William M., "The Wechsler Intelligence Scale for Children: Review of a Decade of Research," Psychological Bulletin, LVII (March, 1960), 132-156.

16. Rainey, Robert G., "A Study of Four School-Ability Tests," The Journal of Experimental Education, XXXIII, No. 4 (Summer, 1965), 306-314.

17. Schachter, Frances F. and Virginia Apgar, "Comparison of Preschool Stanford-Binet and School WISC IQs," Journal of Educational Psychology, XLIX, No. 6 (December, 1958), 320-323.

18. Sheldon, William D. and George Manolakes, "A Comparison of the Stanford-Binet, Revised Form L, and the California Test of Mental Maturity (S-Form)," Journal of Educational Psychology, XLV (December, 1954), 499-504.

19. Stanley, Julian C., "Review of the CTMM S-F, 1963 Revision," The Sixth Mental Measurements Yearbook, edited by Oscard K. Buros, Highland Park, New Jersey, The Gryphon Press, 1965.

20. Tatham, Clifford B. and Arthur A. Dole, "A Note on the Relationship of CTMM S-F to the Revised Binet, Form L-M," Journal of Clinical Psychology, XIX (July, 1963), 302.

21. Wechsler, David, Wechsler Intelligence Scale for Children, New York, Psychological Corporation, 1949.

22. Wechsler, David, The Measurement and Appraisal of Adult Intelligence, Baltimore, Williams and Wilkins Company, 1958.

CHAPTER II

METHOD

Subjects and Procedures for Collecting Data

The subjects (Ss) used in this study were 120 school children who had been referred to the educational clinic of a large school district in the Southwest. The children were referred for problems in the areas of learning, behavior, and mental retardation. As it was desired to compare the CTMM and the WISC for different subgroups of the clinical population, these children were divided into four groups.

Group I was composed of 30 Caucasian children who were neither already in special education classes nor recommended for such classes subsequent to testing at the clinic. Ages in this group ranged from seven to fifteen, and grades encompassed were the second through the eighth. These children were in regular classes.

Group II was composed of 30 Caucasian children who were either already in or were placed in Minimally Brain Injured (MBI) special education classes subsequent to the evaluation. In order for a child to qualify for MBI

placement in this school system, a recent neurological examination is required, the results of which are positive. Ages in this group ranged from seven to twelve and grades included were the second through the fifth.

Group III was composed of 30 Caucasian children who were either already in or were placed in Opportunity Classes subsequent to the evaluation. In order for a child to qualify for Opportunity Class placement in this school district, he must test in a 50-70 IQ range on an individual test of intelligence such as the WISC. Ages in this group ranged from seven to fifteen, encompassing grades two through eight.

Group IV was composed of 30 Negro children who met the same requirements as Group III. The subjects were selected by drawing folders consecutively from the files, starting with the most recent and working backwards, and all cases were accepted which met the following criteria:

(1) That complete CTMM IQ scores were reported, i.e., Language, Non-Language, and Total Mental Factor IQs.

(2) That the S had been administered a WISC and that scores were complete for Verbal, Performance, and Full Scale IQs.

(3) That the two tests had been administered within a year of each other.

(4) That the S fulfilled the requirements for placement in one of the above groups.

Consecutively filed cases at the educational clinic have only temporal proximity in common as the filing system is based on date of referral. It was assumed that the samples of cases selected were representative of the clinical populations for this study.

The CTMM S-F, 1963 Revision, Levels 1, 2, 2H, and 3 had been routinely administered as part of the school testing program. All WISCs had been administered by persons employed as school psychologists by the school system.

Procedures for Treating Data

The data were treated statistically as follows:

(1) Pearson product-moment correlations were calculated to determine the degree of relationship between CTMM and WISC IQ scores between Language CTMM and Verbal WISC IQs, Non-Language CTMM and Performance WISC IQs, and Total CTMM and Full Scale WISC IQs for each of the four groups. The significance of these

correlations was determined by reference to an appropriate statistical table provided by Underwood (3, p. 231).

(2) The test of homogeneity for more than two values of r provided by Edwards (1, pp. 83-84) was employed to determine whether the correlations differed significantly for the four samples.

(3) A *t* test for correlated means was used to determine the significance of mean differences relevant to the second hypothesis. A table provided by McNemar (2, p. 430) was consulted for the statistical values of *t*.

(4) Means and standard deviations of the two tests were calculated for each group.

(5) The .05 probability level ($P = .05$) was chosen as the lower limit for determining the significance of statistical results.

CHAPTER BIBLIOGRAPHY

1. Edwards, Allen L., Experimental Design in Psychological Research, New York, Rinehart & Company, Inc., 1960.

2. McNemar, Quinn, Psychological Statistics, New York, John Wiley and Sons, Inc., 1963.

3. Underwood, Benton J. and others, Elementary Statistics, New York, Appleton-Century-Crofts, Inc., 1954.

CHAPTER III

RESULTS AND DISCUSSION

In Table I can be found the means and standard deviations for all measures for the four clinical groups.

TABLE I

MEANS AND STANDARD DEVIATIONS OF THE CTMM AND WISC FOR THE FOUR GROUPS

Group	CTMM Variable	CTMM Means	CTMM SDs	WISC Variable	WISC Means	WISC SDs
I*	Language IQ	100.13	17.39	Verbal IQ	95.87	13.27
(N=30)	Non-Lang. IQ	100.13	14.66	Perf. IQ	98.53	14.23
	Total IQ	100.43	16.53	Full Scale IQ	96.80	13.72
II	Language IQ	96.23	11.13	Verbal IQ	92.30	10.73
(N=30)	Non-Lang. IQ	95.37	12.51	Perf. IQ	90.77	12.62
	Total IQ	95.07	11.32	Full Scale IQ	90.50	10.92
III	Language IQ	68.67	14.55	Verbal IQ	66.53	5.58
(N=30)	Non-Lang. IQ	69.67	12.76	Perf. IQ	69.77	7.77
	Total IQ	66.30	13.02	Full Scale IQ	64.90	5.22
IV	Language IQ	69.80	9.80	Verbal IQ	65.27	4.70
(N=30)	Non-Lang. IQ	67.30	12.72	Perf. IQ	63.27	9.80
	Total IQ	65.27	10.27	Full Scale IQ	60.93	5.87

*Group I refers to the Heterogeneous Group, Group II the Minimally Brain Injured Group, Group III the White Retarded Group, and Group IV to the Negro Retarded Group.

As can be seen, the means of the CTMM were higher for all variables in all groups with the exception of the CTMM Non-Language and WISC Performance Scale variables in the White Retarded Group. The standard deviations of the CTMM were also larger in all instances but one: the CTMM Non-Language standard deviation in the MBI Group was slightly smaller than the WISC Performance Scale standard deviation for this group.

The generally higher means of the CTMM found here are in agreement with data reported by Littell (3) concerning the research on the comparability of the WISC and the Binet. Since the Binet was the criterion instrument for the CTMM, by inference it might be deduced that lower IQs would generally be found on the WISC than on the CTMM, especially in groups I and II of this study. Using this type reasoning, however, the generally lower WISC IQs found in the retarded groups of this study are in contradiction to evidence reported by Littell. The results found by Altus (1) are also in contradiction to present results, as Altus found virtually no differences between the CTMM S-F and the WISC for either means or standard deviations. It should be recalled, however, that the CTMM is not the Binet, nor was the CTMM S-F used by Altus the same revision as used

in this study. Further, the populations used in this study differed from those of previous studies.

The smaller standard deviations obtained on the WISC indicate the distribution of WISC IQs was more leptokurtic than that of the CTMM. Although the smaller standard deviations found on the WISC perhaps result in more stability of measurement, it could also be argued that it was a less sensitive instrument. It is doubtful, however, that much support could be found among psychologists for the latter portion of the above statement.

Further results and discussion of this study will be presented in the same order as the formal hypotheses. It will be recalled that the first hypothesis was that significant positive correlations would be found between the CTMM and the WISC for subscale and full scale IQs for each group. Data relating to this hypothesis are shown in Table II. It presents the correlations for the CTMM Language and WISC Verbal Scale IQs, CTMM Non-Language and WISC Performance Scale IQs, and CTMM Total and WISC Full Scale IQs for each group.

TABLE II

CORRELATIONS BETWEEN THE CTMM AND THE WISC

Group	CTMM Variable	WISC Variable	r	P
I	Language	Verbal	.78	.01
(N=30)	Non-Language	Performance	.68	.01
	Total	Full Scale	.79	.01
II	Language	Verbal	.33	.05
(N=30)	Non-Language	Performance	.58	.01
	Total	Full Scale	.54	.01
III	Language	Verbal	.32	.10
(N=30)	Non-Language	Performance	.39	.05
	Total	Full Scale	.54	.01
IV	Language	Verbal	-.02	NS
(N=30)	Non-Language	Performance	.05	NS
	Total	Full Scale	.00	NS

As is evident, with the exceptions of the correlations for the CTMM Language and WISC Verbal Scale in the White Retarded Group and all variables in the Negro Retarded Group, correlations were significant at the .05 level or better. Thus, the first hypothesis was partially disconfirmed. Significant correlations between the CTMM and the WISC were

found to exist in the Heterogeneous Group and the Minimally Brain injured Group for each variable considered in this study. Significant correlations also obtained between the CTMM Non-Language and WISC Performance Scale, and the CTMM Total and WISC Full Scale IQs for the White Retarded Group. From a total of twelve correlations, eight were statistically significant and four were non-significant.

The correlations for the Heterogeneous Group were in substantial agreement with the results obtained by Altus (1), while the correlations in the other groups were not. The attenuated range in the other groups possibly accounts for the lower correlations therein obtained. There are other noteworthy aspects of the data from Table II:

(1) the correlation for the CTMM Non-Language and WISC Performance variable was higher than that for the Language-Verbal or Total-Full Scale variables in the MBI Group, which is interesting as the descending order of correlations is usually total, verbal, and non-verbal for composite groups;

(2) the correlations between the retarded groups differed markedly, which would indicate that race may have been an important differentiating factor as the range, ages, and sex variables were comparable.

The four non-significant correlations would indicate the two tests are not interchangeable for those variables in the groups showing such non-significant correlations. As anticipated, merely using non-significant correlations as a criterion for determining the non-interchangeability of IQs for the two tests produced only moderate evidence to support an assumption of non-interchangeability. The rationale for hypothesizing that significant correlations would obtain was based on the observation that major IQ tests generally tend to correlate significantly with each other when given to the same population. It should be noted, too, that even though a correlation is statistically significant the error involved in estimating one score from another can still be appreciable. This is partially because, with large samples, small correlations can be statistically significant.

The second hypothesis of this study was that although significant correlations might be found between the two instruments, the correlations across the groups for the variables in this study would differ significantly. If such were found to be the case, the correlations between the CTMM and WISC would be inconsistent from one group to another. Table III presents data bearing on

this hypothesis. The correlations from the fourth group were omitted from this test since they were not significantly different from zero.

TABLE III

TEST OF HOMOGENEITY FOR THREE VALUES OF r FOR GROUPS I, II, AND III

Variables	χ^2	P
CTMM Language and WISC Verbal Scale	8.478	.02
CTMM Non-Language and WISC Performance	4.330	.20 NS
CTMM Total and WISC Full Scale	3.826	.20 NS

The results shown in Table III fail to support completely the second hypothesis of this study. Evidence was found, nevertheless, which indicated that the correlations were not consistent with respect to CTMM Language and WISC Verbal Scale IQs. Such a statement could not be made concerning CTMM Non-Language and WISC Performance, and CTMM Total and WISC Full Scale IQs, however, since chi squares for these variables only approached significance at the .20 level. Thus, although the second hypothesis was technically disconfirmed, there was some significant evidence indicating that all four correlations for each

variable could not have come from the same population. Such evidence suggests that the correlations between the two tests may vary across differing groups.

The third hypothesis of this study held that an analysis of the mean differences in IQ scores between the two tests would indicate the tests were not equivalent for subscale and full scale IQs for each group. The rationale for using this statistical technique involved an assumption concerning both means and correlations in testing interchangeability. For instance, scores of two tests might correlate highly but the score values could be much higher on one test than on the other. Additionally, the mean scores of two tests could be the same although performance on one test might have little relationship with performance on the other test. Therefore, in order to demonstrate equivalence or non-equivalence of IQ scores on two tests, it is necessary to use both correlations and mean differences.

Data have already been presented bearing on the correlations between the two tests under consideration. A statistical measure of value for comparing two tests with respect to mean differences is the t test for correlated means, and Table IV presents data utilizing this technique.

TABLE IV

MEAN DIFFERENCES, STANDARD DEVIATIONS OF THE DIFFERENCE, AND *t* VALUES FOR THE FOUR GROUPS

Group	Variables	Mean Differences	SDdiff	*t*	P
I	CTMM LIQ and WISC VSIQ	4.27	10.92	2.10	.05
(N=30)	CTMM NLIQ and WISC PSIQ	1.60	11.59	.74	NS
	CTMM Ttl. and WISC FSIQ	3.63	10.23	1.91	.10
II	CTMM LIQ and WISC VSIQ	3.93	12.28	1.72	.10
(N=30)	CTMM NLIQ and WISC PSIQ	4.60	11.53	2.15	.05
	CTMM Ttl. and WISC FSIQ	4.57	10.71	2.30	.05
III	CTMM LIQ and WISC VSIQ	2.13	13.80	.83	NS
(N=30)	CTMM NLIQ and WISC PSIQ	-.10	12.12	-.04	NS
	CTMM Ttl. and WISC FSIQ	1.40	11.08	.68	NS
IV	CTMM LIQ and WISC VSIQ	4.53	10.94	2.23	.05
(N=30)	CTMM NLIQ and WISC PSIQ	4.03	15.66	1.39	.20
	CTMM Ttl. and WISC FSIQ	4.33	11.83	1.97	.10

The data in Table IV show that the mean differences were significant for four of the twelve possibilities; thus the third hypothesis is partially disconfirmed. The mean differences were significant for the following variables

and groups: (1) the CTMM Language and WISC Verbal IQs in the Negro Retarded Group; (2) the CTMM Non-Language and WISC Performance, CTMM Language and WISC Verbal IQs for the MBI Group; and (3) the CTMM Language and WISC Verbal IQs for the Heterogeneous Group. Non-significant differences characterized the White Retarded Group, where none were significant. Excluding the White Retarded Group, differences approaching significance were found for all other variables except one: the CTMM Non-Language and WISC Performance IQs for the Heterogeneous Group.

Of particular interest, it is noted that the difference between means and correlational methods of determining non-interchangeability of IQ scores for the two instruments did not overlap appreciably. Four correlations in Table II indicated non-interchangeability of scores, and four values indicated such non-equivalency. Of these eight combined indicators of non-interchangeability of IQs, there was agreement on only one variable. Thus the value of using both statistical methods for studies of this type is emphasized.

An additional point that can be made from the data in Table IV is that assuming a normal distribution of differences, IQ differences as large as or larger than the

absolute values of the standard deviations of the difference could be predicted for 32 per cent of the subjects. Likewise, score differences between the two instruments with absolute values twice as large as the standard deviation of the difference could be predicted for 5 per cent of the subjects, etc. Information of this type would be an useful adjunct to data that are customarily published in test manuals. It will be recalled that one of the points made in discussing the significance of the study was that IQ test disparities between tests might be interpreted more realistically if test users were familiar with data bearing on the size discrepancies that could be expected to occur. Bradshaw (2) emphasized this point in his study concerning the stability of the CTMM-SF.

Thus far, statistical results have been presented describing the populations under consideration, and which relate to the three hypotheses of this study. All hypotheses were partially disconfirmed, although significant evidence of non-interchangeability of IQs between the CTMM S-F, 1963 Revision and the WISC was found. More specifically, it was found that

(1) both the means and standard deviations of the CTMM were larger than those of the WISC for eleven of the twelve possibilities;

(2) eight of the twelve correlations between the two tests were significant at the .05 level or better, and that the two tests correlated relatively highly with each other in the Heterogeneous groups (r = .68 to .79);

(3) the combined statistical measures used to identify non-interchangeability of IQ scores between the two instruments indicated the IQ scores were non-interchangeable for seven of the twelve comparisons if the criteria of non-significant correlations and significant mean differences are employed;

(4) race appeared to be a significant differentiating factor in the retarded groups, as none of the variables were interchangeable in the Negro Retarded Group.

Therefore, one might conclude that for the populations considered in this study there was little evidence to support an assumption of interchangeability of IQ scores between the CTMM S-F, 1963 Revision and the WISC. It will be recalled

that this finding is in contradiction to that of Altus who concluded that "within a comparable school referral setting, the WISC and CTMM are markedly comparable as to group assessment and roughly comparable as to individual scores" (1, p. 144). A further inference is that it is not possible to predict WISC IQs from CTMM S-F IQs with any appreciable degree of certainty in this school-clinical settings, as it is not usually known which "group" classification a pupil might fit into prior to individual testing. This would appear to be so even for the variables which this study did not identify as being non-interchangeable. It should be pointed out, however, that even though correlations were not high within most of these restricted range groups, particularly the mentally retarded ones, this does not mean that both tests might not place approximately the same children in these groups. Within the groups, though, estimating one IQ from another would not be done accurately.

It is not known to what extent the IQ score differences found in this study were due to the effects of time, faulty test administration, lack of investment, and other extraneous considerations. For practical purposes, however, significant evidence was found indicating that IQs from the CTMM S-F and the WISC were, largely, not inter-changeable for these

particular school-clinical populations. Therefore, it would seem useful to be aware that relatively large IQ differences might be expected to occur in comparable settings. Again, however, the few studies comparing these two instruments do not warrant any but the most tentative conclusions. The fact that the existing studies contradict each other to a considerable extent largely precludes generalizations, and emphasizes the need for further studies comparing these two tests.

CHAPTER BIBLIOGRAPHY

1. Altus, Grace T., "Relationships Between Verbal and Non-Verbal Parts of the CTMM and WISC," Journal of Consulting Psychology, XIX (April, 1955), 143-144.

2. Bradshaw, Donald H., "Stablity of California Test of Mental Maturity IQ's from the Second to the Fourth Grade," Educaticnal and Psychological Measurement, XXIX, No. 4 (Winter, 1964), 939-939.

3. Littell, William M., "The Wechsler Intelligence Scale for Children: Review of a Decade of Research," Psychological Bulletin, LVII (March, 1960), 132-156.

CHAPTER IV

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

This study was done to determine whether IQ scores for the California Short-Form Tests of Mental Maturity, 1963 Revision, and the Wechsler Intelligence Scale for Children were interchangeable in a school-clinical setting. Case files of 120 school children referred to the educational clinic of a large Southwestern school district were used for the raw data, and the subjects were drawn in such a manner as to include four separate clinical groups: (1) a Heterogeneous Group; (2) a Minimally Brain Injured Group; (3) a White Retarded Group; and (4) a Negro Retarded Group. It was assumed that all the groups were atypical of the normative ones for the two instruments, and that by using separate groups, differences between the two instruments might be further delineated.

It was hypothesized that the scores for the two tests were non-interchangeable, and although not formally hypothesized, it was also assumed that both mean differences and correlations are necessary to demonstrate interchangeability of IQ scores. Both the manner of stating hypotheses

and the statistical treatment reflected this assumption. Briefly, the hypotheses held that significant correlations would obtain between the two instruments for all groups and variables, but that further statistical treatment of the data would indicate the IQ scores were, nonetheless, non-interchangeable for all groups and variables.

The statistical methods involved the use of simple correlational coefficients, a test of the homogeneity of r, and a test of mean differences. The significance of statistical results was determined by the use of t tests for correlations and the mean differences, and chi squares were used to determine the significance of the tests of homogeneity of r's. The .05 probability level (P = .05) was chosen as the lower limit for determining significance of statistical findings. The variables considered were the CTMM Language and WISC Verbal Scale IQs, CTMM Non-Language and WISC Performance Scale IQs, CTMM Non-Language and WISC Performance Scale IQs, and CTMM Total and WISC Full Scale IQs for each group.

The findings of the study failed to completely support the hypotheses as exceptions to the hypotheses were found in each instance. Nevertheless, the data substantially supported the assumptions that the IQ scores for the two tests were non-interchangeable for the majority of the

populations and variables considered in this study, and that mean differences are a necessary adjunct to correlations in demonstrating interchangeability or non-interchangeability of IQ scores.

Specifically, of the twelve comparisons considered, only the CTMM Non-Language and WISC Performance, CTMM Total and WISC Full Scale IQs for the Heterogeneous and White Retarded Groups appeared to be even approximately interchangeable. A further finding emerging from the separation of the groups was that race appeared to be an important factor affecting the relationship between the two sets of scores in the Negro Retarded Group. It was found that IQ scores for all the variables were non-interchangeable in this group. Regarding the statistical measures used, four comparisons were identified as non-interchangeable by the correlational method and four were so identified by a test of the significance of mean differences; there was overlap on only one comparison. The test of homogeneity of r indicated the correlations were not comparable across the group. It was concluded that the scores between the two tests were largely non-interchangeable for the populations considered in this study.

Although it was not known to what extent the IQ score differences found in this study were due to the effects of time, faulty test administration, lack of investment, and other extraneous considerations, it would be useful to know if such differences might realistically be expected to occur in comparable school-clinical settings. It was recommended that further studies of this type be conducted, both because existing studies contradict each other, and because the paucity of studies comparing these two widely used tests do not warrant any but tentative conclusions regarding the relationship of the two tests to each other.

BIBLIOGRAPHY

Books

Buros, Oscar K., The Sixth Mental Measurements Yearbook, Highland Park, New Jersey, The Gryphon Press, 1965.

Burstein, Alvin G., "Review of the WISC," The Sixth Mental Measurements Yearbook, edited by O. K. Buros, Highland Park, New Jersey, The Gryphon Press, 1965.

Cronbach, Lee J., Essentials of Psychological Testing, New York, Evanston, and London, Harper and Row, 1960.

Edwards, Allen L., Experimental Design in Psychological Research, New York, Rinehart & Company, Inc., 1960.

Fraser, Elizabeth D., "Review of the WISC," The Sixth Mental Measurements Yearbook, edited by O. K. Buros, Highland Park, New Jersey, The Gryphon Press, 1965.

Freeman, Frank S., Theory and Practice of Psychological Testing, New York, Holt, Rinehart and Winston, 1963.

McNemar, Quinn, Psychological Statistics, New York, John Wiley and Sons, Inc., 1963.

Stanley, Julian C., "Review of the CTMM S-F, 1963 Revision," The Sixth Mental Measurements Yearbook, edited by O. K. Buros, Highland Park, New Jersey, The Gryphon Press, 1965.

Underwood, Benton J. and others, Elementary Statistics, New York, Appleton-Century-Crofts, Inc., 1954.

Wechsler, David, The Measurement and Appraisal of Adult Intelligence, Baltimore, Williams and Wilkins Company, 1958.

Articles

Altus, Grace T., "A Note on the Validity of the Wechsler Intelligence Scale for Children," Journal of Consulting Psychology, XVI (June, 1952), 231.

_____, "Relationships Between Verbal and Non-Verbal Parts of the CTMM and WISC," Journal of Consulting Psychology, XIX (April, 1955), 143-144.

Arnold, Frank C. and Winifred K. Wagner, "A Comparison of Wechsler Children's Scale and Stanford-Binet Scores for Eight- and Nine-Year Olds," Journal of Experimental Education, XXIV (September, 1955), 91-94.

Bradshaw, Donald H., "Stability of California Test of Mental Maturity IQ's from the Second to the Fourth Grade," Educational and Psychological Measurement, XXIV, No. 4 (Winter, 1964), 935-939.

Estes, Betsy W., Mary E. Curtin, Robert A. DeBurger, and Charlotte Denny, "Relationships Between 1960 Stanford-Binet, 1937 Stanford-Binet, WISC, Raven, and Draw-A-Man," Journal of Consulting Psychology, XXV (October, 1961), 388-391.

Finley, Carmen J., Jack M. Thompson, and Albert Cognata, "Stability of the California Short-Form Test of Mental Maturity: Grades 3, 5, and 7," California Journal of Educational Research, XVII, No. 4 (September, 1966), 157-168.

Harlow, Justin E., Jr., Arthur C. Price, Louise J. Tatham, and Jack F. Davidson, "Preliminary Study of Comparisons Between Wechsler Intelligence Scale for Children and Form L of Revised Stanford-Binet Scale at Three Age Levels," Journal of Clinical Psychology, XIII (January, 1957), 72-73.

Littell, William M., "The Wechsler Intelligence Scale for Children: Review of a Decade of Research," Psychological Bulletin, LVII (March, 1960), 132-156.

Rainey, Robert G., "A Study of Four School-Ability Tests," The Journal of Experimental Education, XXXIII, No. 4 (Summer, 1965), 306-314.

Schachter, Frances F. and Virginia Apgar, "Comparison of Preschool Stanford-Binet and School WISC IQs," Journal of Educational Psychology, XLIX, No. 6 (December, 1958), 320-323.

Sheldon, William D. and George Manolakes, "A Comparison of the Stanford-Binet, Revised Form L, and the California Test of Mental Maturity (S-Form)," Journal of Educational Psychology, XLV (December, 1954), 499-504.

Tatham, Clifford B., and Arthur A. Dole, "A Note on the Relationship of CTMM S-F to the Revised Binet, Form L-M," Journal of Clinical Psychology, XIX (July, 1963), 302.

Reports

California Test Bureau Personnel, Technical Report on the California Test of Mental Maturity Series, 1963 Revision, Monterey, California, California Test Bureau, 1965.

Manuals

Clark, Willis W., et. al., California Short-Form Test of Mental Maturity; Manual (Level 2, 1963 Revision), Monterey, California, California Test Bureau, 1963.

Wechsler, David, Wechsler Intelligence Scale for Children, New York, Psychological Corporation, 1949.