THE VALIDITY OF THE LEITER INTERNATIONAL PERFORMANCE SCALE IN MEASURING THE INTELLIGENCE OF SELECTED SUPERIOR CHILDREN

by

EARL FREDERICK PEISNER

A THESIS

submitted to

OREGON STATE COLLEGE

in partial fulfillment of
the requirements for the
degree of

DOCTOR OF EDUCATION

June 1956

APPROVED:

Redacted for privacy

Associate Professor of Psychology

In Charge of Major

Redacted for privacy

Head of School of Education

Redacted for privacy

Chairman of School Graduate Committee

Redacted for privacy

Dean of Graduate School

Date thesis is presented April 26, 1956

Typed by Lesta Peisner

ACKNOWLEDGMENT

To Dr. William R. Crooks, the writer owes a debt of gratitude for his guidance and suggestions during the course of the study and the preparation of the thesis.

The writer also extends his gratitude to the other members of his graduate committee for their help in putting the thesis into final form.

The cooperation of Dr. Wendell Van Loan, Superintendent of the Corvallis Public Schools, Mrs. Charline Edwards, Principal of Harding School, and the teachers of Harding School, was greatly appreciated.

And finally, acknowledgment is made of the willingness and enthusiasm with which the sixth grade pupils of Harding School entered into the test administration.

TABLE OF CONTENTS

TABLE OF TABLES

THE VALIDITY OF THE LEITER INTERNATIONAL PERFORMANCE SCALE IN MEASURING THE INTELLIGENCE OF SELECTED SUPERIOR CHILDREN

CHAPTER I

INTRODUCTION

For the past half-century, since Binet and Simon constructed their intelligence scale to aid in the identification of feebleminded school children, there have occurred numerous attempts to expand and refine measuring instruments which yield information about intelligence. Practical demands from education, the military, and business and industry for objective, reliable and valid measurement of intelligence have been instrumental in the construction of intelligence scales of various kinds. The use of measures of intelligence in describing, predicting, and to some extent controlling human behavior has been an important factor in the selection, placement, and general guidance of individuals. Nevertheless, there exists a constant need for improving measuring instruments so that decisions involving intelligence can be more accurately effected.

Recent efforts have included both verbal and performance scales, as well as combinations of both methods in the same instrument. These developments have come about in an attempt to gain as complete a measure of intelligence as possible, for it has been discovered in

several studies that performance scales measure intellectual factors not assessed by the verbal-type intelligence scales. (4, p. 164) and 7, p. 202)

The use of performance scales has come about for various reasons. They were originally constructed as a substitute for or supplemental to verbal scales such as the Stanford-Binet Intelligence Scale. However, they have assumed special importance in measuring intelligence in situations where the verbal-type instrument is not appropriate; to examine the deaf, the illiterate, non-English-speaking subjects, individuals with reading difficulties, and in general in situations where the subject is likely to be handicapped by verbal tests. Some performance scales have been especially constructed to minimize cultural influences on test performance, and have been referred to as culture-free scales. Further, subjects who might be unable to demonstrate their real skills on a verbal scale because of emotional factors could be more easily identified and a more accurate measure obtained with the use of a performance scale. Performance scales have proved especially useful through the opportunity afforded for clinical observation of the subject. Freeman writes in this regard:

> Clinical psychologists are agreed that, where indicated, the use of performance scales can provide more information than just a rating

> in the form of a numerical index. These tests provide an opportunity to observe qualitative aspects of behavior under standardized conditions in a variety of situations. A subject's approach to a problem might reveal, for example, a state of depression or agitation; hesitation or impetuousness; thoughtful deliberateness, bull-headed persistence, or easy discouragement; an insightful approach or one of haphazard trial-and-error. (7, p. 218)

Although correlations between the Stanford-Binet Intelligence Scale and performance scales are positive, they are low enough to suggest that a performance scale is not interchangeable with this type of intelligence scale. (4, p. 164) and 7, p. 202) Nevertheless, the need still exists, and attempts are periodically made to devise performance scales which measure functions of intelligence comparable to those measured by verbal-type instruments such as the Stanford-Binet.

THE LEITER INTERNATIONAL PERFORMANCE SCALE

The Leiter International Performance Scale is an instrument designed to measure functions of intelligence comparable to those measured by verbal-type scales in situations where the verbal-type scale is inappropriate. The first scale, devised in 1927, was an attempt to measure intelligence by memory and rate of learning. It included one test painted on a fourteen-notch frame. Feebleminded children were the subjects of these early

experiments. During the next two years, eleven additional tests were constructed and a point system for scoring them was devised. These were incorporated into the 1929 scale. A 1930 scale was constructed, including forty-four new tests, and standardized on public school children in Honolulu.

Subsequent scales were constructed in 1936 and 1938, the former containing eighteen tests, the latter, fifty-six. In the 1936 scale the point-scoring method was replaced by a mental-age system. Considerable research was conducted in an effort to determine the reliability and validity of these scales in measuring the intelligence of various racial groups.

The 1940 scale, with sixty-eight tests ranging from age two through age eighteen, was the result of the relocation of the tests of the 1938 scale, and the addition of new tests which were constructed to fill the gaps left by this relocation as well as for use at the odd year levels above year ten. This revision was constructed so as to parallel as closely as possible the 1937 Revision of the Stanford-Binet Intelligence Scale in organization and scoring.

The latest revision, and the one used in this study, was published in 1948. On the basis of experience with

the 1940 scale with high school students and army personnel, only tests at the even-year levels were included beyond year ten. In addition, several test changes were made in order to make the 1948 Revision interchangeable with the Arthur Adaptation of the Leiter International Performance Scale through the twelve year level. (13, pp. 1-57)

Below are listed the tests from year six through year eighteen. None are noted below year six since that level was the lowest reached in establishment of the basal age for subjects in this study. More complete information about the entire scale and directions for administration can be found in Part II of the Manual For The 1948 Revision of The Leiter International Performance Scale. (14)

YEAR VI

(4 tests, 3 months each)

1. Analogous progression
2. Pattern completion test
3. Matching on a basis of use
4. Block design

YEAR VII

(4 tests, 3 months each)

1. Reconstruction
2. Circle series
3. Circumference series
4. Recognition of age differences

YEAR VIII

(4 tests, 3 months each)

1. Matching shades of gray
2. Form discrimination
3. Judging mass
4. Series of radii

YEAR IX

(4 tests, 3 months each)

1. Dot estimation
2. Analogous designs
3. Block designs
4. Line completion

YEAR X

(4 tests, 3 months each)

1. Foot print recognition
2. Block design
3. Concealed cubes
4. Block design

YEAR XII

(4 tests, 6 months each)

1. Block design
2. Similarities; two things
3. Recognition of facial expressions
4. Classification of animals

YEAR XIV

(4 tests, 6 months each)

1. Concealed cubes
2. Analogous designs
3. Memory for a series
4. Form completion

YEAR XVI

(4 tests, 6 months each)

1. Code for a number series
2. Reversed clocks
3. Dot estimation
4. Block design

YEAR XVIII

(6 tests, 6 months each)

1. Position analogy
2. Dot estimation
3. Form completion
4. Concealed cubes
5. Spatial orientation
6. Concealed cubes

The figure on page 8 represents the frame and blocks used in the 1948 Revision of the Leiter International Performance Scale. The test reproduced in the figure is Similarities; two things at the twelve year level. The materials are arranged as indicated in the figure, and the subject's task is to place the blocks in the stalls so that each block is correctly matched with the design appearing on the cardboard strip attached to the frame. There is no time limit on this test.

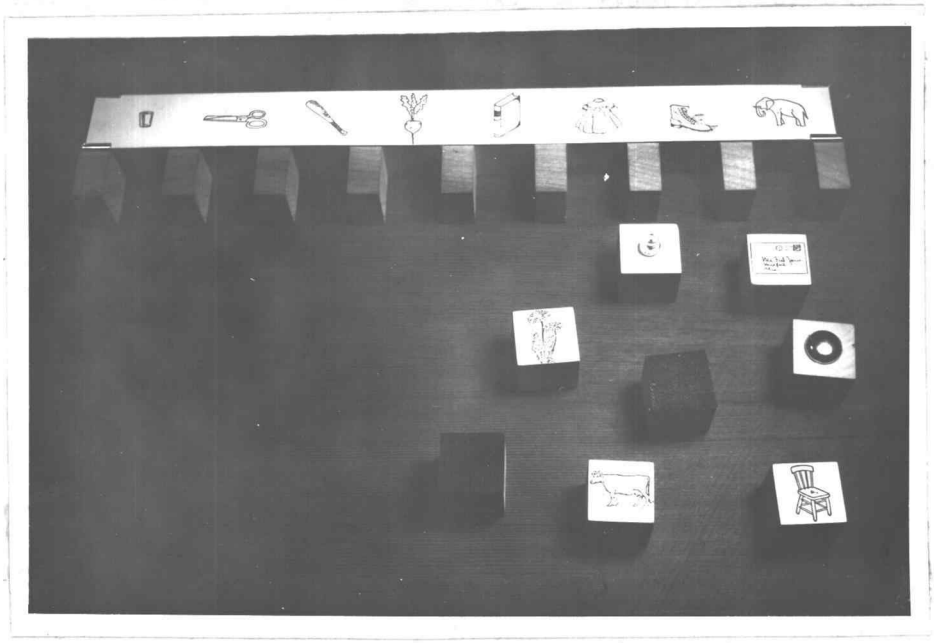

Figure 1. Test XII-2. Similarities; two things.

PURPOSE OF THE STUDY

The purpose of this study was to determine the validity of the Leiter International Performance Scale in measuring the intelligence of selected superior children. The method used in attempting to determine validity was by comparing scores earned on the Leiter International Performance Scale to those of the Stanford-Binet Intelligence Scale and Wechsler Intelligence Scale for Children. Through this contribution it was hoped that the general validity of the Leiter International Performance Scale would be more accurately determined so that it could be effectively used in appropriate situations for the measurement of intelligence.

During the fifty years which have elapsed since the introduction of the Stanford-Binet scale, increased emphasis has been directed toward the problems of the normal and gifted. This trend has been a factor, to some extent, in the emphasis in measurement. In recent years educators have been seeking early and thorough identification of superior children.

The verbal-type intelligence scale has been and still is one of the most effective techniques used to identify the superior child. (24, p. 14) However, the performance-type intelligence scale holds promise of providing additional information which should yield a more thorough and accurate intellectual appraisal. Anastasi writes in this regard:

> On the other hand, the "verbalist" type of individual may obtain a deceptively high score on certain verbal tests, although his understanding of most problems may be very superficial and his practical judgment may be seriously deficient. It is now generally recognized that performance or non-language tests are not simply a substitute for verbal tests. Each type of test predicts somewhat different criteria. Together they provide a more complete picture of the individual and serve as mutual correctives in the evaluation of his test performance. (1, p. 236)

Attempts to assess the intelligence of children who give evidence of superior intellectual achievement poses problems peculiar to the superior group. Paul Witty,

writing in "School and Society" states:

> If by gifted children we mean those youngsters who give promise of creativity of a high order, it is doubtful if the typical intelligence test is suitable for use in identifying them. For creativity posits originality, and originality implies successful management, control, and organization of new materials.....The content of the intelligence test is patently lacking in situations which disclose originality or creativity. (23, p. 504)

The Leiter International Performance Scale, with its unique method and novel test items, and through its attempt to minimize previous learnings, would appear to require of an individual more management, control, and organization of new materials than do most verbal-type intelligence scales. (23, p. 504)

DEFINITIONS

Intelligence is comprehensively defined by Stoddard: ".....ability to undertake actions that are characterized by (1) difficulty, (2) complexity, (3) abstractness, (4) economy, (5) adaptiveness, (6) social values, (7) the emergence of originals, and to maintain such actions under conditions that demand a concentration of energy and a resistance to emotional forces." (19, p. 4) This definition would seem to include the somewhat diverse definitions of Terman, Wechsler and Leiter. Terman defines intelligence as "the ability to carry on abstract

thinking;" (11, p. 123) Wechsler states that "intelligence is the aggregate or global capacity to act purposefully, to think rationally, and to deal effectively with his environment;" (22, p. 3) Leiter implies that intelligence is the ability of the individual to adapt himself to his environment. (13, p. 68)

The superior child is defined here according to the Merrill classification, which refers to individuals whose intelligent quotients are 120 or above as superior or very superior. (17, p. 650) This corresponds in terms of intelligent quotient, to the definition of intellectually gifted, according to the Educational Policies Commission of the National Education Association. (5, p. 43)

A performance scale is defined as a series of intelligence test items requiring the physical manipulation of concrete materials rather than verbal responses.

The t-test is a statistical method, used in this study to test the hypothesis that the mean scores of two intelligence scales are equal. The resulting t-value determines, along with the level of significance, the acceptance or rejection of the hypothesis.

Analysis of variance is a statistical method, used in this study to test the hypothesis that the mean scores

of three intelligence tests are equal. The resulting F-value determines, along with the level of significance, the acceptance or rejection of the hypothesis.

Level of significance is a statistical concept best defined as the probability of rejecting a true hypothesis. The level of significance used in this study was five per cent. This means that five per cent of all possible samples will lead to the erroneous rejection of a true hypothesis. It actually indicates the probability that a true hypothesis will be rejected on the basis of a single random sample. Results beyond the one per cent level are also reported when appropriate.

SCOPE OF THE STUDY

The sample selected for this study included thirty-five sixth grade pupils from the Harding elementary school of the Corvallis Public School System. The subjects ranged in age from eleven years one month to twelve years four months. These pupils were selected according to the criterion set forth in the definition of superior children. This particular school was chosen from the three elementary schools in Corvallis because of the special gifted sixth grade group in operation there.

METHOD OF STUDY

Twenty-eight of the sample of thirty-five pupils were selected on the basis of the 1937 Revision Stanford-Binet Intelligence Scale (hereafter referred to as the Stanford-Binet) scores earned in test administrations prior to this study. The remaining eight subjects were secured by administering Stanford-Binets to sixth grade pupils whose achievement and various psychological test scores suggested that they might earn Stanford-Binet intelligent quotients (hereafter referred to as I. Q.) of 120 or above. After the sample of thirty-five had been obtained, the Wechsler Intelligence Scale for Children (hereafter referred to as the Wechsler) and the Leiter International Performance Scale (hereafter referred to as the Leiter) were administered to each pupil. All administrations of the Wechsler and the Leiter were carried out by the same examiner.

The subjects were first informed in a group by the examiner of the purpose of the testing. Each subject again was acquainted with the reasons for testing at the first administration. The examiner was careful to give the same explanation and instructions to each subject. All pupils who had not been administered the Stanford-Binet were given that intelligence scale. Then all

pupils were given individually the Leiter, after which all pupils were administered the Wechsler. This order of test administration was chosen primarily to separate the two verbal-type scales by the performance scale so that the likelihood of practice effects would be reduced. The testing was completed within a six month period.

In addition to the administration of intelligence scales, the subjects had been given standardized achievement tests five months before the first Leiter scale was administered.

The data from these test administrations were treated with various statistical methods to determine the validity of the Leiter in measuring the intelligence of superior children. First, the analysis of variance was employed to test the hypothesis that the mean scores of the three intelligence tests were equal. An analysis of variance was also utilized to test the hypothesis that the mean scores of the Stanford-Binet, Wechsler and Leiter plus 5 were equal. Leiter, in a recent publication, states that until a full scale revision can be made, "whenever psychological examiners wish to compare the results of the 1948 Scale with the results of other tests they may do so very conveniently by adding five points to the I. Q. obtained from the application of the 1948 Revision." (13, p. 58)

Second, since the analysis of variance test indicated significant differences among the three intelligence scales, t tests were worked out for the following hypotheses: (a) the mean scores of the Wechsler and the Stanford-Binet were equal, (b) the mean scores of the Wechsler and the Leiter were equal, (c) the mean scores of the Stanford-Binet and the Leiter were equal, (d) the mean scores of the Wechsler verbal scale and the Leiter were equal, (e) the mean scores of the Wechsler performance scale and the Leiter were equal, (f) the mean scores of the Stanford-Binet and the Wechsler verbal scale were equal, and (g) the mean scores of the Stanford-Binet and the Wechsler performance scale were equal. The same hypotheses were also tested between the Wechsler scales and the Leiter plus 5 and between the Stanford-Binet and the Leiter plus 5. The five per cent level of significance was used in all tests of hypotheses.

Third, correlation coefficients were computed among the three intelligence tests, as well as between each intelligence test and the standardized achievement test scores earned by the subjects. Part of the sample had taken the elementary form of the California Achievement Tests while some of the subjects were administered the advanced form. Raw score equivalents were computed for the scores earned on the two forms.

LIMITATIONS OF THE STUDY

Perhaps the most obvious and serious limitation is that the validity of the Leiter International Performance Scale was determined by comparison with other instruments, so that errors inherent in these instruments will be projected into the validation data.

A second possible limitation is the size of the sample, which may reduce to some extent the meaningfulness of the tests of significance and the correlational results.

Third, the group involved in the study is not representative in terms of intelligence or socio-economic status, thereby limiting the application of findings to comparable groups.

A final important limitation to be noted is that one of the criteria representing intelligence in this study is scholastic achievement, as measured by a standardized achievement battery. This criterion possibly limits the scope of the concept of intelligence as defined by Stoddard. (19, p. 4)

CHAPTER II

SURVEY OF LITERATURE

Although the primary purpose of this study was to determine the validity of the Leiter International Performance Scale in measuring the intelligence of superior children, information pertinent to this problem and applicable to verbal-type appraisal of intelligence was obtained about the Stanford-Binet and the Wechsler Intelligence Scale for Children. Since these instruments were essential in testing the hypotheses in this study, as well as being important as scales for testing intelligence, selected studies are presented which are similar to some of the hypotheses stated involving verbal-type scales. Further, several studies are noted relative to the application of some of the more widely used performance scales, with which the Leiter can be compared. These studies are summarized below.

THE LEITER INTERNATIONAL PERFORMANCE SCALE

Only three of the approximate twenty-five studies pertaining to the Leiter seem to be relevant to the present investigation. By far the majority of these studies have involved subjects who would be handicapped by a

typical verbal-type intelligence test; non-English speaking subjects, mental defectives, subjects with speech and hearing disorders, and cerebral palsy cases. For example, Beverly and Bensberg obtained a correlation coefficient of .62 between the Stanford-Binet and the Leiter for fifty mental defectives ranging from six years eleven months to sixteen years two months. (2, pp. 89-91) Further, many of the studies employed the 1938 or earlier Leiter scales, and data from these investigations would not seem appropriate here since the 1948 Revision included numerous changes. (13, pp. 28, 57) However, one investigation involving the 1940 scale is included, since Leiter refers to a study of 180 unselected subjects between eight years no months and sixteen years eleven months in which a correlation of .92 was obtained between the 1940 and 1948 Leiter revisions, (13, pp. 57-58) indicating close similarity between the two scales.

Glenn compared the results of the application of the Leiter and the Stanford-Binet to fifty-three unselected children between the ages of six years and six years eleven months. A correlation coefficient of .77 was found between the two arrays of scores. Of the five subjects with Stanford-Binet I. Q.'s above 115, only one tested in the superior range on the Leiter, while the

remaining four earned scores in the normal range. Glenn concluded that the Leiter seems to measure consistently low, as compared to the Stanford-Binet, in the below-average and average ranges but is unpredictable in the above-average ranges. Note Table I below. (8, pp. 20-22)

TABLE I

MEANS, INTELLIGENCE QUOTIENTS, AND STANDARD DEVIATIONS OF THE 1948 LEITER SCALE AND THE STANFORD-BINET

	Mean I. Q.	Mean Standard Deviation
Stanford-Binet	93.36	16.59
1948 Leiter Scale	87.70	19.96

Tate, in attempting to determine how culture-free the Leiter was, administered to 108 children, five years of age, the Leiter, Stanford-Binet, and the Arthur Point Scale, Form I. She found correlation coefficients of .81 and .80 between the Leiter and Stanford-Binet and the Leiter and Arthur Point Scale respectively. A coefficient of .75 was found between the Stanford-Binet and the Arthur Point Scale. Further data from Tate's study can be had from Table II. (20, pp. 497-501)

TABLE II

RANGES, MEANS, AND STANDARD DEVIATIONS OF THE STANFORD-BINET, LEITER, AND ARTHUR POINT SCALE, FORM I

	Range	Mean	Standard Deviation
Stanford-Binet	69-166	114.34	18.8
Arthur Point Scale	58-172	112.15	20.4
Leiter	58-136	93.76	19.1

Williams applied the 1940 scale and the Stanford-Binet to fifty children equally distributed between the ages of six years no months and ten years eleven months. A correlation coefficient of .67 was obtained between the 1940 Leiter Scale and the Stanford-Binet. Additional data are included in Table III. (13, pp. 52-53)

TABLE III

MEANS, INTELLIGENCE QUOTIENTS, AND STANDARD DEVIATIONS OF THE 1940 LEITER SCALE AND THE STANFORD-BINET

	Mean I. Q.	Mean Standard Deviation
Stanford-Binet	107.4	13.44
1940 Leiter Scale	95.6	14.83

COMPARISON OF VERBAL-TYPE AND PERFORMANCE SCALES

MacMurray's study, comparing gifted and dull-normal children with the Pintner-Paterson Scale and the

Stanford-Binet (1916), revealed pertinent information about the performance of gifted children. A correlation coefficient of .23 was found for fifty gifted subjects between the Pintner-Paterson Scale and the Stanford-Binet. These subjects ranged in Stanford-Binet I. Q.'s from 120 to 189, and in chronological age from seven years nine months to ten years seven months. (15, pp. 273-280)

Hamilton tested forty subjects, ranging in age from six years to twelve years eleven months, with the Stanford-Binet and the Revised Arthur Performance Test, Form II. (9, pp. 44-49) Table IV includes data from Hamilton's study related to this discussion.

TABLE IV

RANGES, MEANS, AND STANDARD DEVIATIONS OF THE STANFORD-BINET AND REVISED ARTHUR PERFORMANCE TEST, FORM II

	Range	Mean	Standard Deviation
Stanford-Binet	74-166	108.0	20.5
Revised Arthur	72-141	100.0	16.6

Hamilton further stated that all subjects with Stanford-Binet I. Q.'s above 115 earned lower I. Q.'s on the Revised Arthur. The algebraic average of I. Q. differences for subjects with Stanford-Binet I. Q.'s from 126

to 135 was -20.0, and for subjects with Stanford-Binet I. Q.'s above 136 was -26.6.

Cohen and Collier studied the relationship among the Stanford-Binet, Wechsler, and Revised Arthur Performance Test, Form II, with fifty normal subjects ranging in age from six to eight years. The results of this investigation are given in Table V. (3, pp. 226-227)

TABLE V

CORRELATION COEFFICIENTS, MEANS, AND STANDARD DEVIATIONS OF THE STANFORD-BINET, WECHSLER, AND REVISED ARTHUR PERFORMANCE TEST, FORM II

	Stanford-Binet	Revised Arthur	Wechsler
Wechsler	.85	.80*	
Verbal scale	.82	.77*	
Performance scale	.80	.81*	
Mean	104.8	94.7	99.8
Standard Deviation	15.1	16.4	14.6

*etas

In this study the Wechsler correlated almost as highly with the Revised Arthur Performance Test, Form II, as it did with the Stanford-Binet, and the Wechsler performance scale correlated nearly as high with the Stanford-Binet as did the Wechsler full scale.

Freeman states, regarding comparisons of verbal-type

and performance-type intelligence scales, that when age is held constant, or very nearly so, the correlation coefficients between results obtained with the Pintner-Paterson and similar performance scales, on the one hand, and verbal instruments on the other, drop to between .40 and .60. (7, p. 202)

McBrearty compared the Wechsler and the Arthur Performance Scale, Form I, in relation to the Progressive Achievement Test. His subjects included fifty-two fifth grade children from age ten years three months to twelve years eleven months, with a mean age of eleven years two months. Stanford-Binet I. Q.'s ranged from 50 to 129, with four subjects earning I. Q.'s over 120. Tables VI and VII include pertinent data from McBrearty's study. (16, pp. 15-16)

TABLE VI

CORRELATION COEFFICIENTS AMONG THE ARTHUR PERFORMANCE SCALE, FORM I, THE WECHSLER SCALES, AND THE PROGRESSIVE ACHIEVEMENT TEST

	Arthur		Wechsler verbal		Wechsler Perf.		Wechsler Full	
Wechsler V	.55	.07*						
Wechsler P	.65	.05	.45	.08				
Wechsler FS	.71	.05	.86	.02	.84	.03		
Prog. Achiev.	.56	.07	.81	.03	.50	.07	.78	.04

*Probable error

TABLE VII

RANGES, MEANS, AND STANDARD DEVIATIONS OF THE WECHSLER, ARTHUR PERFORMANCE SCALE, FORM I, AND THE PROGRESSIVE ACHIEVEMENT TEST

	Range	Mean I. Q.	Standard Deviation
Arthur	70-183	101.79	18.63
Wechsler V	62-126	95.83	13.19
Wechsler P	64-132	99.00	13.85
Wechsler FS	71-124	97.12	12.66
Prog. Achiev.	72-122	102.46	13.10

In McBrearty's study the verbal and full scales of the Wechsler correlated much higher with the Progressive Achievement Test than did the Arthur, while the Arthur and Wechsler performance scale correlations with the achievement measure were quite comparable.

VERBAL-TYPE INTELLIGENCE SCALES

Since 1949, when the Wechsler Intelligence Scale for Children was published, numerous studies have been reported in the literature comparing scores earned by subjects on the Stanford-Binet and the Wechsler. (10, p. 152) The three studies which are particularly pertinent to this investigation include one by Krugman, et al, (12, pp. 475-483) which compared the Stanford-Binet and the Wechsler scales in relation to achievement

test scores earned by the subjects, one by Frandsen and Higginson (6, pp. 236-238) which involved subjects with Stanford-Binet I. Q.'s over 120, and a third by Mussen (18, pp. 410-411) dealing with comparisons between the Wechsler and Stanford-Binet, on the one hand, and the Metropolitan Achievement Test on the other.

In a study by Krugman, et al, 332 subjects were administered the Stanford-Binet and the Wechsler. Thirty-seven subjects had a chronological age of eleven years to eleven years eleven months, corresponding very closely in age to the subjects tested in the writers validation of the Leiter. The correlations between the Stanford-Binet and the various part scores of the Wechsler were as follows: verbal scale, .69, performance scale, .53, full scale, .76. Further data of interest obtained in the study cited are noted in Table VIII. Seventy-four subjects earned Stanford-Binet I. Q.'s over 120. For those whose Stanford-Binet I. Q.'s were between 120 and 129 there was a mean difference of plus 19.3 points between their Stanford-Binet score and their Wechsler score, the former being the highest. A mean difference of plus 10.5 points in favor of the Stanford-Binet was noted for subjects whose Stanford-Binet I. Q.'s were over 130. (12, pp. 475-483)

TABLE VIII

MEANS, STANDARD DEVIATIONS, AND t-VALUES OF THE STANFORD-BINET AND WECHSLER

	Mean	Standard Deviation	t-value
Wechsler full scale	101.73	10.95	3.92*
Stanford-Binet	108.35	15.45	
Wechsler verbal scale	104.57	11.95	2.02*
Stanford-Binet	108.35	15.45	
Wechsler performance scale	98.49	12.25	4.33*
Stanford-Binet	108.35	15.45	

*Sig. at 1% level

Frandsen and Higginson administered the Stanford-Binet and the Wechsler to fifty-four unselected fourth grade pupils and correlated the results with educational age scores earned on the Stanford Achievement test. The subjects were of average ability and ranged in age from nine years one month to ten years three months. Pertinent data from this study follows in Table IX. Data of particular interest to be noted in Table IX are first, the correlation coefficient of .80 between the Stanford-Binet and the full scale of the Wechsler, second, the difference between the coefficients of the Wechsler full scale and the Stanford-Binet (.76 and .63 respectively) when correlated with the Stanford Achievement Test, and third, the fact that the Wechsler performance scale

TABLE IX

CORRELATION COEFFICIENTS, MEANS, AND STANDARD DEVIATIONS
OF THE STANFORD-BINET, WECHSLER, AND STANFORD ACHIEVEMENT TEST

	Stanford Achievement	Stanford-Binet	Wechsler Full	Wechsler Verbal	Wechsler Performance
Wechsler					
Full scale	.76	.80			
Verbal scale	.62	.71			
Perf. scale	.65	.63	.88	.52	
Stanford-Binet	.63				
Mean	4.56	105.8	102.4	100.9	103.5
Standard Deviation	.77	11.15	11.15	12.25	11.20

correlated as highly with the Stanford Achievement Test as did the Stanford-Binet (.65 and .63 respectively). (6, pp. 236-238)

Mussen, et al, compared the Wechsler and Stanford-Binet as related to Stanford Achievement Tests in arithmetic and reading, with subjects from seven to thirteen years of age. A correlation coefficient of .85 was obtained between the Wechsler and the Stanford-Binet. Results from this study are noted in Table X. (18, pp. 410-411)

TABLE X

CORRELATIONS AMONG THE STANFORD-BINET, WECHSLER, AND STANFORD READING AND ARITHMETIC TESTS

	Stanford Reading	Stanford Arithmetic
Wechsler		
Full scale	.69	.44
Verbal scale	.73	.47
Performance scale	.57	.29
Stanford-Binet	.65	.45

In conclusion, the studies cited above generally indicated high positive correlations between the Stanford-Binet and the Wechsler, usually higher for the full scale and verbal scale of the Wechsler than the performance scale.

Low positive to medium-positive correlations are generally discovered when correlating verbal-type intelligence scales with performance scales, although some exceptions can be noted.

In studies with normal children, correlations between the Leiter and verbal-type intelligence scales tend to be high positive. Indications have been noted that with superior subjects, as determined by the Stanford-Binet, the correlation coefficients between the Leiter and verbal-type scales are lower than with subjects whose scores are in the average range on the Stanford-Binet.

CHAPTER III

FINDINGS

The findings of this study are reported in three areas corresponding to the basic methods used in testing the hypothesis; first, the analysis of variance of the three intelligence scales, second, the computation of correlation coefficients among the intelligence scales, and third, the computation of correlation coefficients between the intelligence scales and the achievement test battery.

MEAN INTELLIGENCE QUOTIENTS OF THE STANFORD-BINET, WECHSLER AND LEITER INTELLIGENCE SCALES

The mean I. Q. scores and the standard deviations of the three intelligence scales are tabulated in Table XI.

Analysis of variance tests provided F-values significant at the five per cent level, indicating significant differences among the mean intelligence quotients of the three intelligence scales. A two-way classification was used in the analysis of variance tests. See Table XII and Table XIII for the F-values and related data.

Since the analysis of variance test indicated significant differences among the mean scores of the three intelligence scales, it was necessary to apply t-tests to

TABLE XI

MEAN INTELLIGENCE QUOTIENTS AND STANDARD DEVIATIONS OF THE STANFORD-BINET, WECHSLER, AND LEITER SCALES

	No.	Mean I. Q.	Standard Deviation
Stanford-Binet	35	139.54	11.87
Wechsler			
Full scale	35	128.00	9.44
Verbal scale	35	130.11	10.20
Performance scale	35	120.46	11.40
Leiter	35	113.74	13.32

determine between which scales differences occurred. In addition, the verbal and performance scales of the Wechsler were compared with the other mean intelligence scale scores. In determining the significance of differences, t-values exceeding +2.00 or -2.00 indicate a significant difference at the five per cent level; t-values exceeding +2.65 or -2.65 indicate a significant difference at the one per cent level. All t-values except that resulting from comparison between the Wechsler performance scale and the Leiter plus 5[1] were significant at the one per cent level; that is, the probability that the

[1]It will be remembered that Leiter plus 5 refers to Leiter's statement suggesting that Leiter I. Q. scores would more likely be comparable to those of other intelligence scales if five I. Q. points were added to the I. Q. score earned on the Leiter.

TABLE XII

RESULTS OF ANALYSIS OF VARIANCE FOR THE STANFORD-BINET, WECHSLER, AND LEITER SCALES

Variation due to	Sums of squares	Degrees of freedom	Mean square	F-value
Scales	11,691.68	2	5845.84	86.14
Pupils	9,717.05	34	285.80	4.21
Error	4,614.32	68	67.86	
Total	26,023.05	104		

TABLE XIII

RESULTS OF ANALYSIS OF VARIANCE FOR THE STANFORD-BINET, WECHSLER, AND LEITER PLUS 5

Variation due to	Sum of squares	Degrees of freedom	Mean square	F-value
Scales	7,601.68	2	3800.84	56.01
Pupils	9,717.05	34	285.80	4.21
Error	4,614.32	68	67.86	
Total	21,933.05	104		

difference in mean intelligence tests scores was due to chance variations in sampling was less than one in one hundred. The difference between the Wechsler performance scale and the Leiter mean intelligence quotients was significant at the five per cent level, while the difference between the Wechsler performance scale and the Leiter plus 5 was not significant at the five per cent level. The t-values are included in Table XIV.

TABLE XIV

t VALUES FOR MEAN INTELLIGENCE QUOTIENTS OF THE STANFORD-BINET, WECHSLER, AND LEITER SCALES

	Difference M_1 and M_2		Leiter + 5
Wechsler f vs Leiter	14.26	5.07*	3.30*
Wechsler v vs Leiter	16.37	5.68*	3.95*
Wechsler p vs Leiter	6.72	2.23**	.57
Binet vs Leiter	25.80	8.43*	6.80*
Binet vs Wechsler f	11.54	4.42*	
Binet vs Wechsler v	9.43	3.51*	
Binet vs Wechsler p	19.08	6.12*	

*significant beyond the 1% level, (2.65)
**significant beyond the 5% level, (2.00)

CORRELATION COEFFICIENTS AMONG THE STANFORD-BINET, WECHSLER, AND LEITER SCALES

A second attempt to determine the validity of the Leiter International Performance Scale was through the computation of Pearson product-moment correlation coefficients among the Stanford-Binet, Wechsler, and Leiter scales. The correlation coefficients, together with the significance of the coefficients, are given in Table XV.

CORRELATION COEFFICIENTS BETWEEN THE INTELLIGENCE SCALES AND THE CALIFORNIA ACHIEVEMENT TEST

The third general method employed to determine the validity of the Leiter International Performance Scale in measuring the intelligence of selected superior children was to compare the correlation of the Leiter and the California Achievement Test with the correlation obtained between first, the Stanford-Binet and the California Achievement Test, and second, the Wechsler and the California Achievement Test. Comparisons were also made of the correlations obtained between the Wechsler and the California and the Stanford-Binet and California. These correlations appear in Table XVI, and the t-values indicating the significance of the differences among the various correlations are provided in Table XVII.

TABLE XV

PEARSON PRODUCT-MOMENT CORRELATION COEFFICIENTS
AMONG THE STANFORD-BINET, WECHSLER, AND LEITER SCALES

	Leiter	Stanford-Binet	Wechsler Performance	Wechsler Verbal
Stanford-Binet	.42 (±.14) S.E.			
Wechsler				
Full scale	.60 (±.11)	.62 (±.11)	.82 (±.06)	.81 (±.06)
Verbal scale	.64 (±.10)	.68 (±.09)	.30 (±.14)	
Performance scale	.55 (±.12)	.61 (±.11)		

Note: All values significant beyond 1% level (.42) except Wechsler verbal scale vs Wechsler performance scale, which is not significant at 5% level.

TABLE XVI

PEARSON PRODUCT-MOMENT CORRELATION COEFFICIENTS BETWEEN THE STANFORD-BINET, WECHSLER, AND LEITER SCALES AND THE CALIFORNIA ACHIEVEMENT TEST

	California Achievement
Stanford Binet	.63 (±.10) S.E.
Wechsler	
Full scale	.43 (±.14)
Verbal scale	.45 (±.14)
Performance scale	.23 (±.16)
Leiter	.20 (±.16)

TABLE XVII

t-VALUES FOR DIFFERENCES AMONG CORRELATION COEFFICIENTS BETWEEN THE STANFORD-BINET, WECHSLER, AND LEITER AND THE CALIFORNIA ACHIEVEMENT TEST

	Wechsler Full	Wechsler Verbal	Wechsler Performance	Stanford-Binet
Stanford-Binet	1.16	1.04	2.04*	
Leiter	1.04	1.16	.12	2.16*

*significant beyond the five per cent level (2.03)

The only significant differences among these correlations, then, were between the Stanford-Binet and Wechsler performance scale and between the Stanford-Binet and the Leiter.

CHAPTER IV

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

SUMMARY

The general hypothesis tested in this study was that the Leiter International Performance Scale is a valid measure of the intelligence of superior children, the criteria of validity being the Stanford-Binet Intelligence Scale, the Wechsler Intelligence Scale for Children, and the prediction of scholastic achievement.

To test this hypothesis, the Leiter International Performance Scale and the Wechsler Intelligence Scale for Children were administered to thirty-five sixth grade children enrolled in the Harding Public School of Corvallis, Oregon. These children were selected on the basis of having intelligence quotient scores of 120 or above on the Stanford-Binet Intelligence Scale. All pupils included in the sample had been administered the California Achievement Test within a year prior to the use of the Leiter International Performance Scale and the Wechsler Intelligence Scale for Children.

Three specific methods were employed to test the basic hypothesis; (a) the mean intelligence quotient scores of the various scales were compared by analysis of variance and t-tests, (b) intercorrelations among the

various intelligence scales were determined, and (c) comparisons between the intelligence scales and the California Achievement Test were made.

The results of the analysis of variance test and the t-tests showed significant differences among the means of the various intelligence scales, with the exception of the Wechsler Performance Scale and the Leiter plus 5. All differences between the mean intelligence quotient scores of the scales were significant beyond the one per cent level with the exception of the comparison made between the Wechsler performance scale and the Leiter; in this case the difference was significant beyond the five per cent level.

Intercorrelations among the intelligence scales indicated moderate positive correlation between the Leiter and Stanford-Binet (.42) and between the Leiter and Wechsler scales (.55 to .64). These correlations were all significant beyond the one per cent level. Even when considering the standard error involved, the correlations between the Leiter and the Wechsler scales were significant beyond the one per cent level.

Comparisons between the various intelligence scales and the California Achievement Test showed significant differences beyond the five per cent level between the Stanford-Binet and Wechsler performance scale and between

the Stanford-Binet and the Leiter. All other differences did not prove significant at the five per cent level, although the correlation coefficient between the Leiter and the California Achievement Test was low positive (.20). The difference between the correlation of the Stanford-Binet and the California and the Leiter and the California was considerable (.63 and .20).

The general hypothesis of this study, that the Leiter is as valid a measure of the intelligence of superior children as is the Stanford-Binet and Wechsler, is rejected on the basis of the results reported above. All scores earned on the intelligence scales and the achievement test are reported in the appendix.

CONCLUSIONS

The primary conclusion to be drawn from the results of this study is that, with samples like the one employed and with criteria comparable to the standardized scales employed, the validity of the Leiter International Performance Scale in measuring the intelligence of superior children would likely be low. This conclusion would still be warranted if, as Leiter suggests, five I. Q. points are added to an individual's I. Q. score earned on the Leiter Scale.

Some related conclusions seem apparent concerning the measurement of the intelligence of the selected superior children in this sample: first, that the Stanford-Binet has higher validity as a predictor of achievement than does either the Wechsler or the Leiter scales, although it should be noted that a validity coefficient of .62, which is "high" as intelligence scales go, indicates a predictive value only approximately twenty per cent better than chance; second, that the Wechsler scale, particularly the Wechsler performance scale, is more nearly equivalent and hence more likely to be interchangeable with the Leiter than is the Stanford-Binet, since the Wechsler performance scale and the Leiter may be measuring common functions of intelligence; and third, that adding five I. Q. points to the I. Q. scores earned on the Leiter would not increase sufficiently the validity of the Leiter in measuring the intelligence of superior children.

Another conclusion is offered on the basis of the writer's observations during the administration of the intelligence scales. Some of the Leiter tests, such as the Form Completion Test, Year XIV, and the Concealed Cubes Test, Year XVIII, because of their difficulty and novelty at the preadolescent level, demand a degree of

adaptiveness and persistence which does not appear to be required on the performance items of the Wechsler.

RECOMMENDATIONS

Some of the following recommendations are based upon the statistical results of this study, while others are the result of the writer's observations. The basis of the recommendations will be made apparent to the reader.

1. The results of the statistical tests employed suggest that the Leiter International Performance Scale should not be used to measure the intelligence of superior sixth grade children when criteria comparable to those employed in this study are used. It is likely that the validity of the scale when used in this way would be low.
2. A point scale should be developed, or a refinement of the mental age method of scoring should be effected so that a subject could earn partial credit for the correct responses he makes to parts of items. For example, in the various Concealed Cubes Tests the subject may discover the principle involved, hit upon an efficient

mathematical technique to arrive at the answers, and respond correctly to seven of the eight parts involved in the item. Yet if he should count seventy-three cubes instead of seventy-two in the eighth part, he receives no credit for the entire item. This appears to be a severe kind of penalty on some items.

3. Further research with the Leiter and its application to the measurement of the intelligence of superior children should be conducted using different criteria than those employed in this study. Additional criteria which might promote the discovery of valuable information about the Leiter include the following:

   (a) The use of a more comprehensive standardized achievement test than the one used in this study.

   (b) A broader concept of intelligence defined as adjustment to new situations, which would include scholastic achievement, but also such factors as social or vocational achievement. Ratings of achievement or

adaptiveness of the pupils by teachers (including grades), parents and peers might well add significance to a broader concept of intelligence, and supply us with a more meaningful criterion.

(c) School subjects of a non-language type; industrial arts, geometry, mechanical drawing, or art.

(d) Comparisons between the Leiter and special aptitude tests of spatial relations, design judgment, and possibly mechanical comprehension.

(e) Comparisons with a factor-type group intelligence test like the Chicago Tests of Primary Abilities, in which the Thurstones have factored abilities like space and reasoning. (21, p. 7)

4. A final recommendation would be that in further studies attempting to determine the validity of the Leiter in measuring the intelligence of superior children the investigator could select the sample of superior children on some basis other than Stanford-Binet intelligence quotients.

BIBLIOGRAPHY

1. Anastasi, Anne. Psychological testing. New York, MacMillan, 1954. 682p.

2. Beverly, Louise and Gerard J. Bensberg. A comparison of the Leiter, the Cornell-Coxe and Stanford-Binet with mental defectives. The American journal of mental deficiency 57:89-91. 1952.

3. Cohen, Bertram D. and Mary E. Collier. A note on the Wechsler Intelligence Scale for Children and other tests of children 6-8 years old. Journal of consulting psychology 16:226-227. 1952.

4. Cronbach, Lee J. Essentials of psychological testing. New York, Harper and Brothers, 1949. 475p.

5. Educational policies commission, National education association of the United States and the American association of school administrators. Education of the gifted. Washington, 1950. 88p.

6. Frandsen, Arden N. and Jay B. Higginson. The Stanford-Binet and the Wechsler Intelligence Scale for Children. Journal of consulting psychology 15:236-238. 1951

7. Freeman, Frank S. Theory and practice of psychological testing. New York, Henry Holt, 1955. 609p.

8. Glenn, Robert T. A comparison of intelligence quotients derived by the Leiter International Performance Scale and the 1937 Stanford Revision of the Binet. Master's thesis. Pittsburgh, University of Pittsburgh, 1951. 42 numb. leaves.

9. Hamilton, Mildred E. A comparison of the Revised Arthur Performance Tests, Form II and the 1937 Binet. Journal of consulting psychology 13: 44-49. 1949

10. Holland, Glen A. A comparison of the WISC and Stanford-Binet I.Q.'s of normal children. Journal of consulting psychology 17:147-152. 1953.

11. Intelligence and its measurement. Journal of educational psychology 12:123-147. 1921.

12. Krugman, Judith et al. Pupil functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children. Journal of consulting psychology 15:475-483. 1951.

13. Leiter, Russell Graydon. Part I of the manual for the 1948 revision of the Leiter International Performance Scale. The psychological service center journal 11:1-68. 1955.

14. Leiter, Russell Graydon. Part II of the manual for the 1948 revision of the Leiter International Performance Scale. The psychological service center journal 2:263-343. 1950.

15. MacMurray, D.A. A comparison of gifted children and of dull-normal children measured by the Pintner-Paterson Scale as against the Stanford-Binet Scale. Journal of psychology 4:273-280. 1937.

16. McBrearty, John F. A comparison of the WISC with the Arthur Performance Scale, Form I, and their relation to the Progressive Achievement Tests. Master's thesis. State College, Pennsylvania State College, 1951. 31 numb. leaves.

17. Merrill, Maud A. The significance of I.Q.'s on the revised Stanford-Binet scales. Journal of educational psychology 29:641-650. 1938.

18. Mussen, Paul, Sanford Dean and Margery Rosenberg. Some further evidence on the validity of the WISC. Journal of consulting psychology 16:410-411. 1952.

19. Stoddard, G.D. The meaning of intelligence. New York, MacMillan, 1943. 504p.

20. Tate, Miriam E. A study of the performance of selected groups of five-year-olds on the Leiter International Performance Scale. The journal of abnormal and social psychology 47:497-501. 1952.

21. Thurstone, L.L. and T.G. Thurstone. The Chicago Tests of Primary Abilities, manual of instructions. Chicago, Science Research Associates, 1943.

22. Wechsler, David. Measurement of adult intelligence. 3d ed. Baltimore, Williams and Wilkins, 1944. 258p.

23. Witty, Paul. Contributions to the I.Q. controversy from the study of superior deviates. School and society 51:504. 1940.

24. Witty, Paul. (ed.) The gifted child. Boston, D.C. Heath, 1951. 338p.

APPENDIX

SCORES RESULTING FROM THE ADMINISTRATION OF THE STANFORD-BINET, WECHSLER AND LEITER INTELLIGENCE SCALES TO THIRTY-FIVE SELECTED SUPERIOR SIXTH GRADE PUPILS

PUPIL	BINET	LEITER	WISC T	WISC V	WISC P	CAT*
1	170	108	138	149	118	268
2	166	117	135	130	133	266
3	154	117	130	133	121	253
4	153	121	136	133	133	277
5	152	116	145	134	149	278
6	150	139	144	150	129	293
7	150	127	131	133	122	228
8	149	123	125	138	107	225
9	149	124	129	133	120	235
10	149	118	123	128	114	161
11	144	118	137	145	121	196
12	144	138	134	138	124	188
13	142	128	138	145	122	261
14	141	110	124	131	111	274
15	140	136	147	143	143	232
16	140	111	114	120	104	259
17	140	112	118	131	100	247
18	139	104	133	137	124	267
19	139	118	133	128	133	260
20	139	79	125	129	115	274
21	138	123	133	133	127	205
22	138	114	132	138	120	164
23	136	102	135	124	140	192
24	136	104	131	126	129	214
25	135	126	120	133	103	218
26	131	112	128	128	124	168
27	129	113	126	130	117	209
28	128	97	112	114	107	178
29	125	109	124	125	118	74
30	125	123	123	126	115	187
31	124	95	117	124	107	161
32	123	104	120	115	121	183
33	122	114	122	114	127	171
34	122	95	112	108	115	205
35	122	86	106	108	103	172
Total	4884	3981	4480	4554	4216	7643
Mean	139.54	113.74	128.00	130.11	120.46	218.37

*Raw score equivalents between the Elementary and Advanced forms of the California Achievement Test were computed by the California Test Bureau, Los Angeles, California.