THE UNIVERSITY OF DETROIT

A COMPARISON OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN AND THE REVISED STANFORD-BINET, FORM L, WITH A GROUP OF SEVEN- AND EIGHT-YEAR-OLD PUBLIC SCHOOL CHILDREN

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

DEPARTMENT OF PSYCHOLOGY

BY

ALVIN LUBBERS

DETROIT, MICHIGAN

AUGUST, 1954

ACKNOWLEDGMENTS

The writer wishes to acknowledge his gratitude and indebtedness to the many persons whose assistance, cooperation, and encouragement have made this study possible. First of all, thanks are extended to Rev. Charles A. Weisgerber, S.J., Ph.D., Chairman of the Psychology Department at the University of Detroit, who laid out the problem in its broader aspects as a foundation for the several related studies of which this is one.

Acknowledgment is also made of the valuable assistance received from Dr. Harry J. Baker, Director of the Psychological Clinic of the Detroit Public Schools, in obtaining permission to administer the tests in the Public School System. Thanks are also expressed to Miss Elizabeth F. McBryde, Principal, Miss Genevieve M. McGreevy, Assistant Principal, Mrs. Hitchcock, School Secretary, and the several teachers of the Angell Elementary School of Detroit for outstanding cooperation and assistance. Finally, thanks are due the pupils of the Angell School, who relinquished the participation in other activities to act as subjects for the tests.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

It is almost five years since the Wechsler Intelligence Scale for Children[1] was first published. Since that time a body of literature dealing with its application to various groups has made its appearance. The value of a test is considerably enhanced by an accumulation of research literature dealing with its accuracy, distinctive features, and dependability. Such studies were generally anticipated by the publishers and psychologists.

It was to be expected that some of these investigations would have to do with the validity of the scale in relation to a number of criteria. One of these criteria is suggested by the concluding statement of a report by Seashore, Wesman, and Doppelt:[2]

> . . . WISC will approximate in meaning (as far as size of the number is concerned) the IQ's secured by the Stanford-Binet Revision.

It is the purpose of the present study to evaluate this

1. Hereafter the Wechsler Intelligence Scale for Children will usually be referred to as the WISC and the Stanford-Binet will occasionally be referred to as the S-B.

2. H. Seashore, A. Wesman, and J. Doppelt, The Standardization of the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1950, 14, 99-110.

statement as applied to seven- and eight-year-old children. The precise problem, stated in terms of statistical procedures, becomes, "Do the mean scores of the Stanford-Binet and the WISC approximate each other, and do the two scales correlate to a degree that permits the successful prediction of the unknown scores of one of the scales from the obtained scores of the other?"

A procedure such as this has been called "circular validation." But, in defense of this approach, it is contended that the Stanford-Binet, by reason of its primacy in the field of individually administered intelligence tests, has come to be regarded as something of a yardstick with which newer tests are compared. No doubt, the most frequent question asked about the WISC has been, "How does it compare with the Binet?" This is likely to be a pertinent question to many workers for some time to come. In order to avoid a circular type of argument, we have to suppose that the Stanford-Binet is of proven validity. Since the proof is by no means overwhelming, any inferences in regard to the superiority or inferiority of either scale must be tempered by this supposition.

It is hoped that this study may contribute to the cumulating data for the evaluation of the WISC and the enhancement of its value.

CHAPTER II

REVIEW OF RELATED STUDIES

Subsequent to the publication of the WISC, studies of the scale, based upon analyses of the standardization data, were reported in the professional literature. The report on the standardization of the WISC, by Seashore, Wesman, and Doppelt,[1] is an expansion of the technical sections of the test manual. It is pertinent to the present study primarily because of the light it throws upon the comparative scores obtained from urban and rural children and the variation to be expected in the scores of different occupational groups.

In this standardization study it was found that urban children of the sample scored higher than rural children by 5.6 IQ points on the WISC Full Scale, by 6 points on the Verbal Scale, and by 4 points on the Performance Scale. Occupational groups of the standardization sample which correspond to the major occupational groups represented in the present study scored above the norm, 100 IQ. The authors conclude their report with a statement which we shall be able to evaluate later in the light of our own findings:

1. H. Seashore, A. Wesman, and J. Doppelt, The Standardization of the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1950, 14, 99-110.

> The reason for setting the standard deviation at 15 is that it approximates the empirical standard deviation of about 16 secured by Terman and Merrill by an age-scale method. With standard deviations so similar, WISC will approximate in meaning (as far as size of the number is concerned) the IQ's secured by the Stanford-Binet Revision.

In another study, Seashore[2] investigated the differences in scores obtained with the WISC Verbal and Performance Scales. This was an exploration of a theoretical basis for score discrepancies as related to socio-economic, urban-rural, and feeble-minded groups. On the basis of the standardization data, Seashore predicts that in only about four per cent of the cases will children have equal IQ's on both scales. He states also:

> About three-fourths of our subjects will show differences of four points or more, one-half of them eight points or more, and one-fourth of them 15 points or more. Five per cent of the children will show V- and P- IQ discrepancies as great as 25 points or more.

Approximately one per cent will show differences of 35 points or more.

Seashore cautions against expecting either a "zero difference or a typical difference between V- and P- IQ's." A difference in these two IQ's may be clinically meaningful; but great importance must not be attached to the mere difference. It must be interpreted in the light of other pertinent information. He suggests that "we need to develop

2. H. Seashore, Differences between verbal and performance IQ's on the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 62-67.

and exercise the habit of always thinking 'plus or minus so much' when we note or discuss a test score."

Regarding score discrepancies as related to various categories composing the standardization sample, he concludes,

> In general summary of these socio-economic analyses, we can say the intra-group IQ discrepancies are much greater and of more significance than inter-group differences in the relative sizes of V- and P- IQ's.

Grove[3] disapproved of the elimination of the Mental Age concept in the WISC and maintained that Wechsler discarded more than he provided an adequate substitute for. Grove regarded this a serious defect in the WISC and offered a method of computing the MA. In 1951, Wechsler[4] also published three methods of arriving at the Mental Age, making it clear, however, that he was not thereby lending support to the concept.

Grove's study suggests two other problems. The first is occasioned by the fact that each child was tested within 1½ months of his mid-year. Some workers have puzzled over the method used to arrive at the tables of Scaled Score Equivalents for various intermediate ages. If linear interpolation was used, it involved an assumption that

3. W. Grove, Mental Age scores for the Wechsler Intelligence Scale for Children. J. clin. Psychol., 1950, 6, 393-397.

4. D. Wechsler, Equivalent Test and Mental Ages for the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 381-384.

development proceeds at the same rate from mid-year to mid-year.

The second problem that Grove presented has to do with crudeness of measurement. He offers an actual test record which, by hypothetically shifting birth dates two days, can be made to yield IQ scores that vary by seven points.

Various validation studies have also been made upon independent samples to compare the IQ's obtained on the Stanford-Binet with those yielded by the three WISC scales. Such studies are based upon samples that tend to be biased in regard to geographic area, urban-rural categories, and social strata; but they represent the results of the WISC in operation "in the field." The essential data from these investigations are summarized in tabular form at the end of this chapter. A brief description of the samples and a statement of the conclusions reached by the workers is given here.

Frandsen and Higginson[5] made a study of the entire fourth grade of one school. The group consisted of 54 subjects of average ability, who ranged in age from nine years, one month to ten years, three months. An analysis was made of the scores obtained with the Stanford-Binet, the WISC, and the Stanford Achievement Test, Intermediate Battery, Form G.

5. A. Frandsen and J. Higginson, The Stanford-Binet and the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 136-138.

It was found that the "IQ norms from the Stanford-Binet and the Wechsler Intelligence Scale for Children are comparable at the average level." The correlation between the Stanford-Binet and the WISC Full Scale (.80) indicates that "to a considerable extent the two tests are measuring the same factor or factors. Of the variance in the Stanford-Binet IQ distribution, 64 per cent is associated with the variance in the WISC distribution." Bearing in mind that intelligence is not the only factor in school achievement, the investigators found that both the Stanford-Binet and the WISC are valid predictors of school achievement.

Krugman, Justman, Wrightstone, and Krugman[6] made a comprehensive study of 332 children in eighteen schools which were representative of varied neighborhoods, socio-economic strata, and ethnic groups in the five boroughs of New York City. The Stanford-Binet and the WISC were administered to ten age groups ranging from $5\frac{1}{2}$ to $15\frac{1}{2}$ years (all within $1\frac{1}{2}$ months of the mid-year). There were an equal number of boys and girls. One hundred seventy one of the records had been used by Wechsler as part of his standardization population.

For all subjects combined, correlations of .817, .739, and .644 were obtained between the Stanford-Binet, Form L, and the WISC Full Scale, Verbal Scale, and Performance

6. Judith Krugman, J. Justman, J. Wrightstone, and M. Krugman, Pupil functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 475-483.

Scale, respectively. Concerning the mean IQ's, the authors conclude:

> The mean IQ's on the Revised Stanford-Binet, Form L, are higher than those on the Wechsler Intelligence Scale for Children. The widest discrepancy is between the Stanford-Binet and the WISC Performance Scale. The WISC Verbal Scale IQ's are closer to the Stanford-Binet (mean difference 5.1 points) than either the WISC Performance (mean difference 10.2 points) or the Full Scale IQ's (mean difference 7.3 points). For the 332 children between $5\frac{1}{2}$ and $15\frac{1}{2}$ years in this study, the mean IQ's were 108.5 on the Stanford-Binet, 103.4 on the WISC Verbal Scale, 98.3 on the WISC Performance Scale, and 101.2 on the WISC Full Scale.

As regards individual scores, approximately two-thirds of the subjects had discrepancies of 10 points or less on the WISC Verbal and Full Scales, as against one-half on the Performance Scale. The greatest differences were associated with higher Stanford-Binet IQ's. Greater differences between the Stanford-Binet and both the WISC Full Scale and Verbal Scale were also found to be associated with lower age levels. This was not true of the Performance Scale.

In their evaluation of the WISC, the authors say, in part:

> In the present study, the WISC gave results similar to the Revised Stanford-Binet, Form L, in the large majority of cases at the lower IQ levels but the discrepancies appearing at the upper IQ levels may be considered too large to permit the use of the WISC in place of the Stanford-Binet until further work has been done. . . . During the period before such data are available and before the evidence is conclusive, psychologists will and should certainly use the WISC, but will still feel the need for checking with the Stanford-Binet in those instances where clinical judgment and WISC results seem at variance.

Kureth, Muhr, and Weisgerber[7] compared the scores obtained by 100 normal children on the WISC and the Revised Stanford-Binet, Form L. Seventy-two of the subjects were institutional children and the rest, with one exception, attended parochial schools in the kindergarten or first-grade. One-half were five-year-olds and the other half were six years of age. The correlation of the Stanford-Binet and the WISC Full Scale for the entire sample was found to be .807. The Stanford-Binet mean score exceeded that of the WISC Full Scale by 8.40 IQ points.

The authors conclude that the correlation is sufficiently high to indicate that the two scales measure about the same thing, but that the WISC yields a lower IQ and, therefore, underestimates the child's intellectual capacity. The discrepancies between individual scores on the two scales ranged from 1 to 28 IQ points. It was also found that there are 3 chances in 5 that a five-year-old child will score at least 10 IQ points lower on the WISC Full Scale than on the Stanford-Binet. The writers regard this as quite serious in clinical work.

Pastovic and Guthrie[8] investigated the relationship of the WISC and the Revised Stanford-Binet, Form L, through the

7. Sister Genevieve Kureth, Jean Muhr, and C. A. Weisgerber, Some data on the validity of the Wechsler Intelligence Scale for Children. Child Develpm., 1952, 23, 281-287.

8. J. Pastovic and G. Guthrie, Some evidence on the validity of the Wechsler Intelligence Scale for Children. J. Consult. Psychol., 1951, 15, 385-386.

analysis of the scores of 50 kindergarten pupils whose mean age was five years and six months and 50 second grade subjects whose mean age was seven years and six months. The data from four Master's Theses and one other unpublished study were also reported. The results contradicted Seashore's statement that the WISC will approximate the IQ's secured by the Stanford-Binet. The writers "conclude that WISC IQ's should not be interpreted as equivalent to a Binet IQ at age levels below ten years since the WISC score is consistently lower than that of the Binet." Rapaport's[9] findings, reported in the above study, agreed closely with those of these writers.

Weider, Noller, and Schramm[10] offer a provisional scale of equivalent scores for the Stanford-Binet, 1937 Revision, Form L, and the WISC Full Scale, based upon the scores obtained from a study of 106 white Louisville children, free from emotional problems. The sample was divided into two sections: a younger group consisting of 44 children whose ages ranged from five years to seven years, eleven months; and an older group of 62 children whose ages ranged from eight years to eleven years, eleven months.

9. I. Rapaport, A comparison of performance on the Wechsler Intelligence Scale for Children and the Revised Stanford-Binet Scale. In J. Pastovic and G. Guthrie, Some evidence on the validity of the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 385-386.

10. A. Weider, P. Noller, and T. Schramm, The Wechsler Intelligence Scale for Children and the Revised Stanford-Binet. J. consult. Psychol., 1951, 15, 330-333.

The correlation between the two scales was unusually high (.89) for the entire group and the mean Stanford-Binet IQ was only 3.1 points higher than the mean WISC Full Scale score. Their table of equivalent scores is based upon the regression equation, y equals $0.85x$ plus 11 (P.E.: 5.8 IQ points), in which y represents the WISC score and x represents the Stanford-Binet score.

It should also be mentioned that there are some reported studies in which the performance of subnormal children upon the two scales are compared. Stacey and Levin,[11] for example, tested 44 morons (WISC Full Scale IQ: 50 to 69) and 26 borderline children (WISC Full Scale IQ: 70 to 81). The correlation of the Stanford-Binet and the WISC Full Scale was .60 for the former group, .44 for the latter, and .68 for the two groups combined. The average age of these children was eleven years and eleven months. Sloan and Schneider[12] tested 40 mental defectives whose ages ranged from nine years and one month to fifteen years and five months. The mean WISC Full Scale IQ for this group was 58.3 and the correlation of the Stanford-Binet with the WISC was .493 for the Full Scale, .751 for the Verbal Scale, and .641

11. C. Stacey and Janice Levin, Correlation analysis of scores of subnormal subjects on the Stanford-Binet and the Wechsler Intelligence Scale for Children. Amer. J. ment. Def., 1951, 55, 590-597.

12. W. Sloan and B. Schneider, A study of the Wechsler Intelligence Scale for Children with mental defectives. Amer. J. ment. Def., 1951, 55, 573-575.

for the Performance Scale. These studies are not included in the tabular summaries at the end of this chapter because the findings of the present study are not applicable to subnormal subjects.

A summary of the results of six validation studies is presented in Tables I and II. Only the data for individual age groups of normal children less than eleven years old have been included. All told there are 621 subjects from five to ten years of age.

As shown in Table I, the correlations reported in the various studies are in rather close agreement. The coefficients for the Stanford-Binet and the WISC Full Scale vary from .710 to .896; for the Stanford-Binet and the Verbal Scale from .630 to .880; and for the Stanford-Binet and the Performance Scale from .486 to .790. In each group, except one, the Stanford-Binet correlated better with the Full Scale than with either of the subscales. It also correlates more highly with the Verbal Scale than with the Performance Scale in ten out of thirteen groups.

Table II presents a comparison of the mean IQ scores yielded by the Stanford-Binet and the three WISC scales for thirteen groups. It will be noted that in every group the mean Stanford-Binet score is higher than any of the three mean WISC quotients. Except in one group, the WISC Full Scale scores fall between the Verbal and the Performance scores in size. As to the interrelationships among the WISC subscales, the Performance Scale mean scores are higher than Verbal scores in six of the thirteen groups.

TABLE I

SUMMARY OF VALIDATION STUDIES SHOWING CORRELATIONS BETWEEN THE STANFORD-BINET AND THE THREE WISC SCALES

Investigators	Mean Age	N	Correlation with criterion: Full Scale	Verbal Scale	Perf. Scale
Frandsen & Higginson	9-7½ (Est.)	54	.80	.71	.63
Krugman, Justman, Wrightstone, & Krugman	5-6	38	.896	.668	.763
	6-6	38	.821	.726	.741
	7-6	43	.733	.642	.486
	8-6	44	.823	.783	.574
	9-6	31	.873	.834	.790
	10-6	29	.856	.880	.538
Kureth, Muhr, & Weisgerber	5-6	50	.844	.790	.727
	6-6	50	.785	.711	.715
Pastovic & Guthrie	5-6	50	.71	.63	.57
	7-6	50	.88	.82	.71
Rapaport	7-6	100	.85	.79	.74
Weider, Noller & Schramm	6-6 (Est.)	44	.90	.82	.79
Total		621			

Note: Although the averaging of coefficients of correlation is not a valid mathematical procedure, the following unweighted means are presented as crude guides: S-B and Full Scale, .829; S-B and Verbal Scale, .754; and S-B and Performance Scale, .675.

TABLE II

SUMMARY OF VALIDATION STUDIES PRESENTING MEAN IQ SCORES OF THE STANFORD-BINET AND THE THREE WISC SCALES

Investigators	Mean Age	Mean Intelligence Quotients S-B IQ	WISC Full Scale	WISC Verbal Scale	WISC Perf. Scale
Frandsen & Higginson	9-7½ (Est.)	105.8	102.4(a)	100.9	103.5
Krugman, Justman, Wright-stone, & Krugman	5-6	107.26	96.08*	97.92*	95.11*
	6-6	111.87	101.87*	103.97*	98.97*
	7-6	105.02	101.07*	103.28(b)	98.40*
	8-6	110.86	104.73*	106.32*	101.55*
	9-6	116.84	106.52*	107.00*	104.42*
	10-6	111.66	105.97*	107.69(c)	103.03*
Kureth, Muhr, & Weisgerber	5-6	102.90	91.70*	89.90*	94.80*
	6-6	107.40	101.80*	98.30*	105.90(d)
Pastovic & Guthrie	5-6	115.08	111.50*	108.56*	112.68
	7-6	113.02	103.16*	101.58*	104.24*
Rapaport	7-6	97.01	89.60*	89.94*	91.60*
Weider et al.	6-6 (Est.)	100.0	92.7	94.8	91.7
Total		1404.72	1309.10	1310.16	1305.90
Unweighted Mean		108.06	100.70	100.78	100.45

* Differences between indicated scores and the Stanford-Binet are significant at the 1 per cent level of confidence. "P" for unmarked scores is not known.

a. t ratio is 0.45 (not significant, statistically).
b. $.30 > P > .20$.
c. $.05 > P > .02$.
d. P equals .280.

CHAPTER III

SUBJECTS AND PROCEDURE

The 48 children who served as subjects in this study were all attending the Angell Elementary School of Detroit, Michigan. The selection of these children came about through a request to Dr. Harry J. Baker, Director of the Psychological Clinic of the Detroit Public Schools, for aid in choosing an "average and typical" group of Detroit school children. The Angell School was designated as representative of the Detroit Public School System.

The school was located in a middle-class neighborhood. The occupations of the children's fathers were mostly along the line of clerical, sales, protective work, and skilled labor. A few were proprietors of small businesses. None were professional workers; a few were factory workers.

In planning the study, it was decided to test only white, American-born children in regular grades who were free from physical defects as well as emotional and behavioral problems. Instead of testing at or near the mid-year, as was done in the standardization samples, it was determined to test at various specific ages throughout the seven- and eight-year levels. It was thought that this procedure would more likely provide an adequate number of subjects and

at the same time, give a basis for the selection of the specific individuals to be tested. The element of chance needed for the selection of a representative sample was introduced by the accident of date of birth.

In the plan of procedure, the youngest subjects, one boy and one girl, were to be 7 years and 15 days of age. Each succeeding pair was to be one month older; that is, they were to be 7 years and 1½ months of age, 7 years and 2½ months of age, etc. Since it was expected that the school could not, in every case, provide a subject of the exact age when the test was to be administered, a 15-day tolerance in either direction was set as the limit of deviation.

The actual selection of subjects was made from the school records. The child who was nearest the criterion age, and met all the other requirements, was tested. In almost every instance, a subject was obtained whose age, at most, was only a few days from the criterion. The mean age of the seven-year-old group was 7 years, 6 months, and ½ day; that of the eight-year-old group was 8 years, 6 months, and ½ day. The grade placement of the subjects is shown in Table III.

The Wechsler Intelligence Scale for Children and the Stanford-Binet, 1937 Revision, Form L, were used throughout. As in the standardization of the WISC, all 12 subtests of this scale were used. All tests were administered by the writer, who followed the instructions in the respective manuals precisely. He was prepared for this work through

courses in individual testing at the graduate level, along with practice testing supervised by his instructors. He had also administered individual tests under supervision as a student clinician, in two well-established and favorably-known clinics.

TABLE III

GRADE PLACEMENT OF SUBJECTS

Group	Second Grade		First Grade	
	First Semester	Second Semester	First Semester	Second Semester
7-year-old boys	4	6	2	0
7-year-old girls	5	7	0	0
8-year-old boys	1	3	5	3
8-year-old girls	0	2	5	5
Total	10	18	12	8

The scales were administered in the order, ABBA; that is, one subject received the Stanford-Binet first, the next received the WISC first, and so on. In no case were both scales administered to a subject on the same day. However, they were administered as closely together as the availability of the subjects and the examiner's schedule permitted. Table IV provides information on the number of days that elapsed between tests.

TABLE IV

TIME INTERVALS FOR TEST ADMINISTRATION

Time Intervals	Number of Subjects
One day	9
Two days	12
Three days	7
Four days	5
Five days	6
Six days	2
Seven days	4
Eight days	1
Nine days	1
Fourteen days	1

It was thought that children of the seven- and eight-year-old levels might vary in efficiency at different times of the day because of fatigue, lagging interest, anticipation of the end of the school day, and other circumstances. For this reason the tests were administered at the same time of day. In no case did the time vary more than a half hour.

The test situation provided through the very fine co-operation of the school authorities, teachers, and other personnel was ideal. A seldom-used, small classroom was assigned to the examiner. This provided a setting to which the child was accustomed, the kind of setting in which he

usually did his school work. Interruptions were a rarity and were mainly of the type that the child accepted as routine; that is, the child accepted fire drills and similar occurrences as part of the school day.

After the first few days, rapport was practically pre-established for the examiner. The children, according to the teachers, regarded it "an honor" to be selected for the tests. The teachers also cooperated by indicating their hearty approval of the testing. None of the children showed reluctance to taking the tests; in fact, the examiner was beseiged in hallways and upon the street by children who asked, "When are you going to take me?"

Only with a few of the youngest subjects did the examiner speak of "playing games" because even the younger subjects disregarded the subterfuge and asked who was going to "correct this test." To the children who seemed to want an explanation, the examiner said that he, too, was "going to school at the University" and that he had "to write a paper about these tests for his instructors." This explanation seemed to satisfy the subjects, and they cooperated fully.

Strict conformity to the plan was costly in time and effort, but the examiner has the satisfaction of believing that the obtained scores are valid indices of the subjects' intelligence as measured by the respective scales.

CHAPTER IV

PRESENTATION AND INTERPRETATION OF DATA

This chapter is a summary of the analysis made of the data through the use of such statistical techniques as are best suited to emphasize the quantitative similarities and differences between the Stanford-Binet and the WISC. Since the study is primarily concerned with the comparison of the quantitative measures, or Intelligence Quotients, that each scale provides, there will be a minimum of analysis of the factors that produced these results, either in the subjects or in the instruments.

In a study of this kind, the interpretation put upon the practical usefulness and meaning of the coefficients of correlation, means, practice effect, regression equation, and other procedures, will be largely influenced by the writer's orientation and the practical use to which the facts are to be put. It is one thing to determine the general trends and differences; it is quite another matter when the instrument is viewed as a measuring scale for determining the individual capacities of a specific patient in the clinical situation. The evaluations found here will lean toward the latter viewpoint.

In the interest of making the results of this study as concise and meaningful as possible, and in conformity with usual practices, the data for each of the two age groups are presented separately and also in combination. The possibility that the material may be found useful in some future research also makes this procedure desirable.

Analysis by Correlation: Full Scale

The coefficient of correlation (r) is a measure of the degree to which two sets of scores vary in relation to one another. It is of value in this study because it expresses, in a single, numerical index, the co-variation of the paired scores that were obtained by the Stanford-Binet and the WISC. The coefficient of correlation is also a measure of the accuracy with which an unknown score can be predicted from a known score.

If the correlation is to result in meaningful coefficients, the data must meet certain requirements, namely, that the variables are logically related, that they are rectilinearily related, and that the sample is representative and adequate in size. The data for this study meet the first two requirements well. The fact that the sample is small and somewhat biased in the direction of superior intelligence is taken into consideration both in the computations and interpretation.

The coefficients obtained by the Pearson product-moment method of correlation are presented in Table V. As will be

seen from this tabulation, the correlation of the Stanford-Binet and the WISC Full Scale is .759 (S.E.: .209) for the seven-year-olds, .638 (S.E.: .209) for the eight-year-olds, and .698 (S.E.: .146) for the seven- and eight-year-olds together. All of these coefficients are positive and significant at the 0.1 per cent level of confidence except that of the eight-year-old group, which is significant at the 1 per cent level.

TABLE V

CORRELATION OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN AND THE STANFORD-BINET, FORM L

WISC Scale	Group	r	σ_r	Sign.
Full Scale	7-year-olds	.759	.209	.1%
	8-year-olds	.638	.209	1%
	7- and 8-year-olds	.698	.146	.1%
Verbal Scale	7-year-olds	.740	.209	.1%
	8-year-olds	.632	.209	1%
	7- and 8-year-olds	.710	.146	.1%
Perf. Scale	7-year-olds	.517	.209	2%
	8-year-olds	.483	.209	5%
	7- and 8-year-olds	.529	.146	.1%

In interpreting these correlations, the following considerations will be found useful. Alternate forms of the same scale, for example, Forms L and M of the Revised

Stanford-Binet, are found to be correlated highly. Terman[1] reports coefficients of approximately .93 for several groups. On the other hand, scales that differ materially in content, structure, and method of scoring show a lower correlation. Wechsler[2] reports that tests, other than the various revisions of the Stanford-Binet, correlate with his Wechsler-Bellevue adult scale from .39 to .81, with a mid-point of about .60. It might then be expected that two scales, both designed primarily to measure the intelligence of children and both similarly standardized, but differing in other features, should have a coefficient of correlation that falls somewhere between the mid-point of the dissimilar scales (.60) and correlation of alternate forms (.93)--possibly about .75. By this criterion, the correlations that we have obtained are satisfactory.

We may also approach the matter empirically. It was noted during our review of related studies in Chapter II, and in the summary of correlations in Table I, that the correlations of the Stanford-Binet with the WISC Full Scale varied from .71 to .90 with a central tendency approaching .83. Of the three separate groups at the seven-year level, two showed r's that were higher than ours (.85 and .88) and one was lower (.73). Among the eight-year-olds, the single

1. Lewis Terman and Maud Merrill, Measuring Intelligence, p. 47.

2. David Wechsler, The Measurement of Adult Intelligence, p. 134.

group reported had an r of .82, which is considerably higher than our r of .64. A difference of this size is not easily explained. It is possible that our sample, being small and easily subject to bias, is not truly representative. This, however, may not be presumed upon the basis of the small amount of evidence now available. Under the circumstances, we are required to interpret the correlation as it stands, rather than arbitrarily reject it as being spurious.

Certain statistical procedures are used for the evaluation of a coefficient of correlation in terms of the efficiency with which the unknown scores of one test can be predicted by the use of the known scores obtained on another test. While the prediction of scores is not of great value by itself, it is the crucial test of the per cent of common factors found in the two scales. If two scales correlate perfectly, having a coefficient of 1.00, the two are interchangeable; they both measure the same thing equally well. But if one test measures a variable to a different degree than the other, a perfect concomitant variation in the scores will not be expected because the individual subjects are almost certain to differ in the degree to which each possesses certain capacities and traits.

The coefficient of correlation is not to be interpreted as a percentage. A better concept of the meaning of a particular r can be obtained from the coefficient of alienation

(k), which is a measure of the absence of relationship.[3] Column 4 of Table VI shows that the k for the correlation coefficients obtained between the Stanford-Binet and the WISC Full Scale are approximately .65 for the seven-year-olds, .77 for the eight-year-olds, and .71 for the two groups combined. Viewed in this way, the coefficients of correlation for the Stanford-Binet and the WISC Full Scale are of only limited value in the prediction of the probable scores of one scale from the obtained scores of the other.

The counter-measure of the coefficient of alienation is known as the index of forecasting efficiency (E). This measure is expressed in terms of the per cent of reduction in errors of prediction when forecasting scores. It is approximately the difference between the coefficient of a perfect correlation (1.00) and the coefficient of alienation (k). Therefore, E is an estimate of the amount of improvement that may be expected in the accuracy of a prediction when the obtained scores of one of the scales are used as a basis for such forecasting. As between the Stanford-Binet and the WISC Full Scale, Column 5 of Table VI shows that these values are approximately 35 per cent for the seven-year-olds, 23 per cent for the eight-year-olds, and 29 per cent for the seven- and eight-year-olds.

3. J. P. Guilford, Psychometric Methods. (New York: McGraw-Hill, 1936.) P. 362. (Hereafter called, "Guilford's Psychometrics.")

TABLE VI

COEFFICIENTS OF ALIENATION, INDICES OF FORECASTING EFFICIENCY, AND THE VALUE OF $100r^2$ FOR THE CORRELATIONS OF THE STANFORD-BINET AND THE THREE WISC SCALES

(1) Scale	(2) Age Group	(3) r	(4) k	(5) E	(6) $100r^2$
WISC Full Scale	A*	.76	.65	35.0**	57.8
	B	.64	.77	23.2	41.0
	C	.70	.71	28.6	49.0
WISC Verbal Scale	A	.74	.67	32.7	54.8
	B	.63	.78	22.3	39.7
	C	.71	.70	29.6	50.4
WISC Performance Scale	A	.52	.85	14.6	27.0
	B	.48	.88	12.3	23.0
	C	.53	.85	15.2	28.1

* Group A: 7-year-olds; Group B: 8-year-olds; Group C: 7- and 8-year-olds.

** This Table overestimates the percentage of reduction in errors slightly (up to 2 per cent), but the percentages given are satisfactory for the point that we wish to make. Interested readers are referred to Guilford's Psychometrics, p. 362, for a formula that will give a more exact estimate.

It will be noted that Column 6 of Table VI is simply headed "100r^2" rather than "coefficient of determination" or "percentage of common causal factors measured," both of which expressions are based on the assumption that scale y contains all the factors in scale x but some others also.[4] This cannot be conceded either upon logical or empirical grounds in the case of the Stanford-Binet and the WISC. The values of 100r^2 are presented in the belief that they are a better measure than r of the actual co-variation of the scores.[5]

The comparisons and statistical measures presented lend themselves to the conclusions that:

1. The coefficients of correlation between the Stanford-Binet and the WISC Full Scale indicate a co-variant relationship which is significantly less than would be expected to exist between alternate forms of the same scale but greater than that found between scales designed to measure different aspects or factors in intelligence.

2. The coefficients of correlation are as large as might be expected on logical grounds and are comparable to the average reported in other studies in the case of our seven-year-old group. The r for our eight-year-old group is .18 lower than that reported for this age group in another study. This is not, however, sufficient reason for

4. Guilford's Psychometrics, p. 305.

5. T. G. Andrews (Ed.), Methods of Psychology. (New York: John Wiley and Sons, 1948.) P. 502.

regarding it as unacceptable.

3. The percentage of error in prediction of the unknown scores is reduced only 23 to 35 per cent through the use of the obtained scores of the other scale.

The practical implications for these findings may be stated as follows:

1. The two scales are not sufficiently alike to be used interchangeably nor are the contributions of each sufficiently unique to justify the inclusion of both in a battery of tests used in the clinic.

2. It is impossible to estimate the scores of one scale from the obtained scores on the other with even a moderate degree of confidence.

3. The Stanford-Binet and the WISC probably measure different factors or elements, or at least measure them in different degrees. Until future research provides a more empirical guide, the clinician will be faced with making the selection, which in his judgment will best serve the clinical purpose.

Analysis by Correlation: Verbal Scale

The Wechsler Intelligence Scale for Children was designed to provide intelligence quotients for the two subscales as well as for the Full Scale. The correlation of each of these subscales and the Stanford-Binet has been computed for the purpose of estimating the presence of common casual factors. The degree to which the two subscales are

testing the same factors or facets of intelligence as the Stanford-Binet can be inferred from their predictive value as expressed in the percentage of reduction of error when the scores of these subscales are used to predict Stanford-Binet scores.

In addition, there is the very practical problem concerning the feasibility of using either one of the subscales separately. While the author and publishers obviously did not intend such use of the scales, those in charge of overloaded but under-staffed clinics will not have been blind to this possibility. The Verbal Scale will be analyzed first.

The point is often made that the Stanford-Binet is heavily loaded with verbal items. For this reason the Verbal Scale might logically be expected to be superior to the Full Scale in correlation with the Stanford-Binet. This is not completely borne out in the reported studies of the scales. Table I which presents a summary of the findings for thirteen groups of normal subjects, shows that in only one of the groups did the Verbal Scale have a higher correlation with the Stanford-Binet than the Full Scale.

In the present study the individual groups followed the same trend, but the differences in the coefficients of correlation were generally smaller. As shown in Table V, the correlation of the Verbal Scale and the Stanford-Binet is .740 (S.E.: .209) for the seven-year-olds, .632 (S.E.: .209) for the eight-year-olds, and .710 (S.E.: .146) for the two groups combined. The WISC Full Scale correlations with the

Stanford-Binet for the corresponding groups were .759, .638, and .698. It is noted, therefore, that the Stanford-Binet correlates better with the WISC Full Scale than with the Verbal Scale in the individual age groups; but when the groups are combined, the reverse is true. In comparing the degree of correlation of the Stanford-Binet with either of these scales in any age group, the difference in the r's is found not to exceed .02. Such small differences do not change the efficiency of prediction, as shown in Column 5 of Table VI, by more than 2.3 per cent.

These findings lead to the conclusions that:

1. The WISC Verbal Scale is slightly less efficient than the Full Scale in predicting Stanford-Binet scores.

2. The WISC Verbal Scale is not an adequate substitute for the Stanford-Binet.

3. The WISC Verbal Scale correlates too well with the Stanford-Binet to be used with it in the same battery of clinical tests. Its contribution would not be sufficiently unique.

Analysis by Correlation: Performance Scale

The WISC Performance Scale may be regarded as an attempt to supply the need of those who believed that the Stanford-Binet was too heavily loaded with verbal items which gave the trained, loquacious, superficially bright child an undue advantage. It has also been contended that the efforts of the constructors of the Binet type of scale

to design a test that would produce a "pure" measure of intelligence has resulted in a restricted scale which does not measure all the important facets and factors of intelligence. The inclusion of a manipulative performance scale in the WISC, standardized upon the same data as the verbal portion, provides an instrument that measures factors of intelligence left unevaluated by the Binet. Not only the type of mental tasks presented to the subject, but the direct scoring for speed and accuracy make radically different demands upon the subject. For these reasons, the Performance Scale of the WISC would not be expected to correlate highly with the Stanford-Binet.

When a comparison is made of the coefficients of correlation of the Stanford-Binet and the WISC Scales, the expected tendencies are found to exist. In the thirteen age groups of normal subjects, summarized in Table I, the Performance Scale correlation with the criterion is inferior to that of the Verbal and Full Scales. Numerically expressed, the average coefficient of the Performance Scale is .079 lower than that of the Verbal Scale and .154 lower than that of the Full Scale.

The coefficients of correlation of the Stanford-Binet and the WISC Performance Scale, as presented in Table V, are .517 (S.E.: .209) for the seven-year-olds; .483 (S.E.: .209) for the eight-year-olds; and .529 (S.E.: .146) for the two groups combined. It should be noted that the level of confidence is 2 per cent for the seven-year-olds and 5 per cent

for the eight-year-olds. They cannot, therefore, be interpreted with the same degree of confidence as the other coefficients. But the r of the combined groups is significant at the 0.1 per cent level. As can be seen from Column 5 of Table VI, the forecasting efficiency of this scale does not exceed 15.2 per cent. It is interesting to note that the combination of the WISC Performance and Verbal IQ's results in a scale that has a higher correlation with the criterion than either of the subscales alone. Individual subtests that show a low correlation with each other and with the criterion are testing a relatively unique aspect of intelligence. When the complete battery provides a sample of all of these unique aspects, then it is a more valid test. In the case of the WISC, the errors in the subscales tend to balance each other out, thus giving the Full Scale a higher correlation than either scale by itself.

This study prompts the following conclusions:

1. The Performance Scale of the WISC, either by reason of the content, the method of scoring, or both, measures, to a considerable extent, the same thing as both the WISC Verbal Scale and the Stanford-Binet; but it also contributes the measure of a unique factor or factors of intelligence.

2. The strictly psychometric contribution of the WISC Performance Scale is, however, probably not sufficiently unique to earn it a place as a separate scale in a battery of clinical tests.

A Study of the Stanford-Binet and the WISC Mean Scores

The analysis made through the technique of correlation provided a method of determining the extent to which the Stanford-Binet and the WISC produced scores that varied concomitantly. A study of the mean scores of the two scales provides a method of comparing the calibration of the scales. In this section, the means of the scores yielded by the Stanford-Binet and the three quotients obtained by the WISC are compared and the significances of their differences interpreted. A comparison will also be made of the mean scores at the higher and lower levels of intelligence to determine whether or not the calibration is equally uniform at both levels.

The data dealing with the mean scores yielded by the Stanford-Binet and the WISC Full Scale are presented in Table VII. It will be noted that the means of the scores yielded by the Stanford-Binet are 117.04 for the seven-year-olds; 111.25 for the eight-year-olds; and 114.15 for the two groups combined. The mean Full Scale scores for the corresponding groups on the WISC are 110.67, 105.42, and 108.04. As is seen from these scores, our sample proved to be considerably above the norm of IQ 100, in spite of an attempt to select a truly representative group. It is recognized that this bias of the sample will preclude the use of these data for drawing inferences about subjects of subnormal intelligence.

TABLE VII

MEAN IQ'S, RANGE, AND SIGNIFICANCE OF THE MEAN DIFFERENCES OF THE STANFORD-BINET AND THE WISC FULL SCALE AT THE SEVEN- AND EIGHT-YEAR LEVELS AND FOR THE TWO GROUPS COMBINED

Age Group	Test	N	Mean	Range	S.D.	$S.E._M$	Diff.	$S.E._D$	t	Sign.
7-year-olds	S-B	24	117.04	49 (91-140)	13.22	2.76	6.37	1.79	3.55	1%
	WISC	24	110.67	38 (93-131)	9.95	2.07				
8-year-olds	S-B	24	111.25	48 (91-139)	9.65	2.01	5.83	1.77	3.29	1%
	WISC	24	105.42	39 (84-123)	10.30	2.15				
7- & 8-year-olds	S-B	48	114.15	49 (91-140)	11.92	1.74	6.11	1.28	4.79	.1%
	WISC	48	108.04	47 (84-131)	10.27	1.50				

It will be noted that for the individual groups the range of the WISC scores is 9 and 11 IQ points less than on the Stanford-Binet, but the difference is only 2 points for the combined groups. The size of the standard deviations, as compared to those of the standardization data of the WISC (15 IQ points) and the Stanford-Binet (16 IQ points),[6] indicates a greater uniformity in our sample. This is probably due to the bias of the sample.

Table VII also shows that the obtained difference in the mean IQ was 6.37 points for the seven-year-olds; 5.83 points for the eight-year-olds; and 6.11 IQ points for the two groups combined. These differences are significant at the 1 per cent level of confidence for the separate age groups and significant at the 0.1 per cent level for the combined groups. It appears, therefore, that the WISC is calibrated in such a way that it yields mean scores that are lower than the Stanford-Binet by about 6 IQ points for children of more than average intelligence, at the seven- and eight-year levels.

This tendency is found rather consistently in other studies, as summarized in Table II. Krugman et al.,[7] for example, found relatively large differences between the

6. Seashore's statement regarding the size of the standard deviations is quoted in Chapter II.

7. Judith I. Krugman, et al., Pupil functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 475-483.

scores of the two scales at the lower age levels. This tendency, which is not so pronounced at the higher age levels, is not easily explained. Muhr,[8] who observed the same trend in her five- and six-year-old subjects, has suggested that the difficulty of the WISC items for children of these ages has the effect of making each test in the series extremely short and, therefore, an unreliable measure of the younger child's true capacity. In our sample, however, because the children were older, this did not appear to be the case. A review of the protocols shows that the subjects "had a try" at a rather wide range of items. Since the tendency of the Stanford-Binet to yield higher scores at the upper level of intelligence is rather consistent, it seems safer to conclude that the discrepancies lie in a difference of calibration in the two scales.

In order to determine whether or not the difference in scores on the WISC and Stanford-Binet could be attributed to either of the two WISC subscales, data similar to those found in Table VII were compiled for these also. Table VIII presents the comparative data for the WISC Verbal Scale, and Table IX shows the corresponding data for the Performance Scale. All differences between the means are significant at the 1 per cent level or better except that between the Stanford-Binet and the WISC Performance Scale for seven-

8. Jean P. Muhr, Validity of the Wechsler Intelligence Scale for Children at the five and six year level. Unpublished Master's Thesis, University of Detroit, 1952.

TABLE VIII

MEAN IQ'S, RANGE, AND SIGNIFICANCE OF THE MEAN DIFFERENCES OF THE STANFORD-BINET AND THE WISC VERBAL SCALE AT THE SEVEN- AND EIGHT-YEAR LEVELS AND FOR THE TWO GROUPS COMBINED

Age Group	Test	N	Mean	Range	S.D.	$S.E._M$	Diff.	$S.E._D$	t	Sign.
7-year-olds	S-B	24	117.04	49 (91-140)	13.22	2.76	8.87	1.85	4.78	.1%
	WISC	24	108.17	38 (91-129)	9.90	2.06				
8-year-olds	S-B	24	111.25	48 (91-139)	9.65	2.01	5.62	1.63	3.44	1%
	WISC	24	105.63	36 (87-123)	8.43	1.76				
7- & 8-year-olds	S-B	48	114.15	49 (91-140)	11.92	1.74	7.25	1.23	5.89	.1%
	WISC	48	106.90	42 (87-129)	9.29	1.36				

TABLE IX

MEAN IQ'S, RANGE, AND SIGNIFICANCE OF THE MEAN DIFFERENCES OF THE STANFORD-BINET AND THE WISC PERFORMANCE SCALE AT THE SEVEN- AND EIGHT-YEAR LEVELS AND FOR THE TWO GROUPS COMBINED

Age Group	Test	N	Mean	Range	S.D.	$S.E._M$	Diff.	$S.E._D$	t	Sign.
7-year-olds	S-B	24	117.04	49 (91-140)	13.22	2.76	5.66	2.50	2.27	5%
	WISC	24	111.38	38 (90-127)	10.72	2.24				
8-year-olds	S-B	24	111.25	48 (91-139)	9.65	2.01	7.21	2.50	2.88	1%
	WISC	24	104.04	53 (79-132)	13.20	2.75				
7- & 8-year-olds	S-B	48	114.15	49 (91-140)	11.92	1.74	6.44	1.73	3.73	1%
	WISC	48	107.71	53 (79-132)	12.48	1.82				

year-olds, which is significant at the 5 per cent level. Discrepancies of WISC IQ from Stanford-Binet IQ are about the same for Full, Verbal, or Performance Scales. Furthermore, the means of the three WISC quotients are so similar that in no case do they differ, for any age group, by more than 3.5 IQ points in our sample.

This line of inquiry may, therefore, be dismissed as an unproductive method. A study of WISC calibration at different levels of intelligence promises to be more rewarding.

Score Discrepancies at Different Levels of Intelligence

An inspection of the raw scores and the results of other studies both indicate that the difference in mean scores yielded by the Stanford-Binet and the WISC Full Scale are greater at the higher levels of intelligence than they are at the average level. To determine the degree of this difference, the means, mean differences, and significances of the differences were computed for both the higher- and lower-scoring groups.

The entire sample of 48 subjects was divided into a "Higher-Scoring Group" consisting of those whose Stanford-Binet IQ scores were 114 or higher and a "Lower-Scoring Group" composed of those children whose Binet IQ scores were below 114. The "Higher-Scoring Group" consisted of 16 seven-year-olds of whom 8 were boys and 8 were girls, and 8 eight-year-olds of whom 3 were boys and 5 were girls. The

"Lower-Scoring Group" consisted of the remaining 24 subjects of the sample.

The data for the "Higher-Scoring Group," as summarized in Table X, show that the mean differences between the Stanford-Binet and the WISC Full Scale scores are 10.06 IQ points for the seven-year-olds; 11.88 for the eight-year-olds; and 10.67 for the two groups combined. All of these differences are significant at the 1 per cent level of confidence or better. While it is impossible to determine at present if the Stanford-Binet is overestimating or the WISC is underestimating intelligence, or if the calibration of both scales is faulty, the discrepancies are sufficiently large to demand the serious attention of anyone using either scale.

The means, differences of the means, and significance of the differences of the "Lower Scoring Group" are shown in Table XI. The mean differences between the Stanford-Binet and the WISC Full Scale are 1.00 IQ point for the seven-year-olds; 2.81 points for the eight-year-olds; and 1.54 IQ points for the seven- and eight-year-olds combined. Only the difference of the eight-year-olds is significant at even the 20 per cent level of confidence. The important features of these differences, so far as interpretation is concerned, are their uniformity and the lack of significance. The differences can be interpreted as due to chance or to sampling error. However, the uniformity of the differences lends itself to the more reasonable possibility that the higher

TABLE X

MEAN IQ'S, RANGE, AND SIGNIFICANCE OF MEAN DIFFERENCES OF THE HIGHER ONE-HALF OF THE STANFORD-BINET SCORES AND THE SCORES OF THE SAME SUBJECTS ON THE WISC FULL SCALE AT THE SEVEN- AND EIGHT-YEAR LEVELS AND FOR BOTH GROUPS COMBINED

Age Group	Test	N	Mean	Range	S.D.	$\sigma_{(M_1-M_2)}$	Diff.	S.E.$_D$	t	Sign.
A*	S-B	16	124.19	26 (114-140)	8.07					
						7.22**	10.06	1.87	5.39	.1%
	WISC	16	114.13	31 (100-131)	8.67					
B	S-B	8	121.75	24 (115-139)	7.21					
						7.54	11.88	2.85	4.17	1%
	WISC	8	109.88	26 (97-123)	8.94					
C	S-B	24	123.38	26 (114-140)	7.88					
						7.38	10.67	1.54	6.93	.1%
	WISC	24	112.71	34 (97-131)	8.99					

* Group A: 7-year-olds; Group B: 8-year-olds; Group C: 7- and 8-year-olds.

** Standard deviation of the mean of the differences; i.e., computed from the distribution of the differences.

TABLE XI

MEAN IQ'S, RANGE, AND SIGNIFICANCE OF MEAN DIFFERENCES OF THE LOWER ONE-HALF OF THE STANFORD-BINET SCORES AND THE SCORES OF THE SAME SUBJECTS ON THE WISC FULL SCALE AT THE SEVEN- AND EIGHT-YEAR LEVELS AND FOR BOTH GROUPS COMBINED

Age Group	Test	N	Mean	Range	S.D.	$\sigma(M_1-M_2)$	Diff.	S.E.$_D$	t	Sign.
A*	S-B	8	102.75	21 (91-112)	8.06					
						4.66**	1.00	1.76	.57	Greater than 20%
	WISC	8	103.75	23 (93-116)	7.55					
B	S-B	16	106.00	22 (91-113)	5.95					
						7.26	2.81	1.87	1.50	20%
	WISC	16	103.19	39 (84-123)	10.13					
C	S-B	24	104.92	22 (91-113)	6.90					
						6.75	1.54	1.41	1.10	Greater than 20%
	WISC	24	103.38	39 (84-123)	9.30					

* Group A: 7-year-olds; Group B: 8-year-olds; and Group C: 7- and 8-year-olds.

** Standard deviation of the mean of the differences; i.e., computed from the distribution of the differences.

Stanford-Binet score of one subject is cancelled out by the higher WISC score of another subject. This is what actually happened in our sample. The important question is whether or not this would happen in other samples, and this we cannot answer. In any case, it is reasonably safe to state the generalization that the mean differences in IQ scores are greatest at the higher levels of intelligence and that the Stanford-Binet usually is the higher of the two.

Individual Discrepancies in Test Scores

The statistical techniques so far used in the comparison of the Stanford-Binet and the WISC are measures of central tendency in which dispersions are regarded as margins of error. For the clinician, however, group tendencies are not of primary importance because he is not working with groups. He is working with individuals and wants to know whether or not the IQ scores obtained for a specific individual can be trusted. But he cannot be sure that his subject's score will not show a large discrepancy.

There is another reason for doubting the dependability of a patient's score. The presence of an individual at a clinic argues in favor of the possibility that the patient is in some way a deviant or an atypical person. That this deviation might well extend to his mental functions is not only possible but probable.

It may be regarded as an inconsistency that the data derived from an especially "normal" sample are presented as

guidance for the clinician. The procedure may be defended on the ground that an understanding of the normal is basic to the detection and evaluation of the deviant. This section is intended to point out the possibilities of wide individual discrepancies in scores, even those of normal subjects. It suggests the use of caution in the interpretation of IQ scores and the need for supplementary clinical evaluation.

Only two of the 48 subjects used in this study received identical Stanford-Binet and WISC Full Scale scores. The number and per cent of cases in which the Stanford-Binet quotients exceeded WISC Full Scale scores by a certain number of IQ points is shown in Table XII, and the corresponding data for cases in which the WISC Full Scale scores were higher are presented in Table XIII. It will be seen that 73 per cent of the subjects had higher scores on the Stanford-Binet and 23 per cent had higher scores on the WISC. Fifty per cent received scores on the Stanford-Binet that were more than five points higher than on the WISC, 33 per cent had scores that were more than ten points higher, and 17 per cent had Stanford-Binet scores that were more than fifteen IQ points higher. In contrast, only 8 per cent had WISC Full Scale scores that were more than five points higher, and 2 per cent had WISC scores that were more than ten IQ points higher than the Stanford-Binet scores.

The patterns of individual score discrepancies at the upper and lower levels of intelligence are presented in

TABLE XII

NUMBER AND PERCENTAGE OF CASES IN WHICH STANFORD-BINET QUOTIENTS EXCEED WISC SCORES BY A CERTAIN NUMBER OF IQ POINTS

Difference in IQ Points	Cases			Cumulative Percentage	
	Number	Cumulative Total	Per Cent of Cases	One Point and Up	25 Points and Less
1	3		6.25		72.9*
3	3	6	6.25	12.5	66.7
4	2	8	4.17	16.7	60.4
5	3	11	6.25	22.9	56.3
6	2	13	4.17	27.1	50.0
7	3	16	6.25	33.3	45.8
9	2	18	4.17	37.5	39.6
10	1	19	2.08	39.6	35.4
11	1	20	2.08	41.7	33.3
12	1	21	2.08	43.8	31.3
13	4	25	8.33	52.1	29.2
14	1	26	2.08	54.2	20.8
15	1	27	2.08	56.2	18.7
16	3	30	6.25	62.5	16.7
17	1	31	2.08	64.6	10.4
18	1	32	2.08	66.7	8.3
20	2	34	4.17	70.8	6.3
25	1	35	2.08	72.9*	

* These percentages are computed on the basis of the entire group of 48 subjects.

TABLE XIII

NUMBER AND PERCENTAGE OF CASES IN WHICH WISC SCORES EXCEED STANFORD-BINET QUOTIENTS BY A CERTAIN NUMBER OF IQ POINTS

Difference in IQ Points	Cases			Cumulative Percentage	
	Number	Cumulative Total	Per Cent of Cases	One Point and Up	25 Points and Less
1	3		6.25		22.9*
3	1	4	2.08	8.3	16.7
4	2	6	4.17	12.5	14.6
5	1	7	2.08	14.6	10.4
7	1	8	2.08	16.7	8.3
8	1	9	2.08	18.7	6.3
10	1	10	2.08	20.8	4.2
11	1	11	2.08	22.9*	

* These percentages are computed on the basis of the entire group of 48 subjects.

Table XIV. Here a comparison is made of the test scores of subjects who made the highest and lowest scores on the Stanford-Binet and the three WISC scales. Of the 6 children who received the highest scores on the Stanford-Binet, 5 were also among the group of 6 subjects who had the highest scores on the WISC Full Scale. All 6 of these children had higher Binet than WISC scores. The mean difference was 12.16 IQ points. Of the 6 children of the sample who received the highest WISC Full Scale scores, 5 had Stanford-Binet scores that exceeded these. For this group, the Binet scores were higher by an average of 6.17 IQ points.

Of the 6 children who had the lowest WISC Full Scale scores (IQ 84 to 95), 4 were also in the group who had the lowest Stanford-Binet scores. Here, too, the Binet scores were higher than the WISC scores in all cases, except one. The average difference is 6.83 IQ points. Among the subjects who had the lowest Stanford-Binet scores (there are 7 in this group because of tie scores), 3 had higher Binet scores and 4 had higher WISC Full Scale scores. For this group, the mean Stanford-Binet scores were higher than those of the WISC by 1.29 IQ points. The writer was unable to discover a pattern or clear-cut trend among the scores of the subscales which would contribute to an understanding of the score discrepancies.

It may be said, then, that the higher scores on both scales tend to be associated with each other and the lower scores are also associated with each other. At both levels

TABLE XIV

A COMPARISON OF TEST SCORES OF SUBJECTS MAKING HIGHEST AND LOWEST STANFORD-BINET AND WISC QUOTIENTS

Scores	Subject Number	S-B IQ	WISC IQ SCORES Full Scale	Verbal Scale	Perf. Scale
Highest	2	140	115	115	111
S-B	46	139	123	118	124
IQ's	19	136	123	119	122
	4	132	131	129	127
	20	132	121	115	124
	8	130	123	119	122
Total		809	736	715	730
Mean		134.83	122.67	119.17	121.67
Highest	4	132	131	129	127
WISC	8	130	123	119	122
IQ's	19	136	123	119	122
	46	139	123	118	124
	42	112	123	110	132
	20	132	121	115	124
Total		781	744	710	751
Mean		130.17	124.00	118.33	125.17
Lowest	33	91	88	89	89
S-B	16	91	98	91	106
IQ's	23	94	95	94	97
	30	99	100	105	94
	41	99	84	87	83
	40	102	92	105	79
	47	102	112	104	120
Total		678	669	675	668
Mean		96.86	95.57	96.43	95.43
Lowest	41	99	84	87	83
WISC	33	91	88	89	89
IQ's	37	105	92	90	96
	40	102	92	105	79
	13	94	93	91	97
	23	94	95	94	97
Total		585	544	556	541
Mean		97.50	90.67	92.67	90.17

the Stanford-Binet yields a higher score, but the difference is greatest at the upper level of intelligence.

The disparity that may occur in extreme cases can be emphasized by pointing out the difference in the classification of a subject and by comparing his rank position on the two scales. This is done for three of the subjects who stood at different levels of intelligence.

Subject No. 2. Age: 7 years.

Stanford-Binet IQ: 140. Classification:[9] Genius or near genius. Rank: above all other 47 subjects.

WISC Full Scale IQ: 115. Classification:[10] Bright Normal. Rank: thirteen subjects had higher scores.

Subject No. 41. Age: 8 years.

Stanford-Binet IQ: 99. Classification: Normal. Rank: three had lower scores and two had equal scores.

WISC IQ: 84. Classification: Dull Normal. Rank: lowest of all subjects.

Subject No. 42. Age: 8 years.

Stanford-Binet IQ: 112. Classification: Superior. Rank: twenty-four subjects had higher scores and four had equal scores.

WISC IQ: 123. Classification: Superior. Rank: one subject had a higher score and three had an equal score.

9. Classification as suggested by Terman, 1916. See C. M. Loutitt, Clinical Psychology. (New York: Harper and Brothers, 1947) p. 97.

10. David Wechsler, Wechsler Intelligence Scale for Children, p. 16.

If these descriptions were listed in random order, without reference to the subjects, it is doubtful if a psychologist would match them correctly. Since test results are often given to persons, other than psychologists, in the form of descriptive phraseology and classification to make them more understandable, the foregoing examples serve to warn anyone making such reports to be very cautious in the phraseology used. He should also be sure to make abundantly clear which scale he is interpreting.

Regression Equation

A regression equation is sometimes used to predict the most probable scores on one test from the known scores on another test. For example, Weider, Noller, and Schramm[11] offer a table of provisional equivalent Stanford-Binet and WISC Full Scale IQ scores based upon the formula, y equals $0.85x$ plus 11 (P.E.: 5.8), in which y represents the WISC score and x is the Stanford-Binet score. This equation was based upon the scores of 106 white children in Louisville, ranging in age from five years to eleven years and eleven months. The coefficient of correlation for the entire group was .89 $\pm$ .02; the mean Stanford-Binet score was 93.1 (S.D.: 19.56) and the mean WISC IQ score was 90.0 (S.D.: 18.90).

11. A. Weider et al., The Wechsler Intelligence Scale for Children and the Stanford-Binet. J. consult. Psychol., 15, 330-333.

With a probable error of 5.8 IQ points, such a table probably has limited practical value.

The regression equation computed from the scores obtained in our study is: y equals .601x plus 39.4 (S.E.: 7.35), in which y represents the WISC Full Scale score and x is the Stanford-Binet score. Since the predictive value of this equation is dependent upon the degree of correlation and the equality of calibration of the two scales, it is not expected that predictions will be accurate. The standard error, 7.35, indicates that an overestimation or underestimation of as much as 22 IQ points may be made in five per cent of the predictions.

As a practical demonstration of the crudeness of the predictions made by this equation, a sample, consisting of every fifth subject was selected and the WISC scores computed from the Stanford-Binet scores. The last three entries are cases deliberately selected for their large score discrepancies. The actual WISC scores and the errors in prediction are shown in Table XV.

From the table, it will be noted that error resulting from the use of our regression equation can be as great as 16 IQ points in an actual case. Since the examiner or clinician who administers an individual test is concerned with the assessment of the intelligence of a specific individual, rather than with group tendencies, the regression equation is almost worthless for predicting scores. He cannot be sure that the predicted score will not be one in which the

error is large.

TABLE XV

ERRORS IN PREDICTION RESULTING FROM THE USE OF A REGRESSION EQUATION

Subject Number	Stanford-Binet Scores	Actual WISC Full Scale Scores	Computed WISC Full Scale Scores	Error
5	112	116	107	9
10	129	111	117	6
15	128	115	116	1
20	132	121	119	2
25	111	119	106	13
30	99	100	99	1
35	105	101	102	1
40	102	92	101	9
45	107	102	104	2
2	140	115	124	9
41	99	84	99	15
42	112	123	107	16

Practice Effect

A small but significant improvement, called practice effect, is sometimes noted in test performance when a second test is administered a few days after the first. Such improvement can be a problem for the clinician because it affects the reliability of the obtained scores. In the present study, practice effect was controlled by the administration of the two scales in counterbalanced order. It is

possible, therefore, to compute the improvement in scores when a test is given second. This was done by dividing the sample into two groups: Group I, consisting of subjects who were given the WISC first; and Group II, composed of those who took the Stanford-Binet first. The means of the four quotients were computed for each group. As shown in Table XVI, the WISC scores of Group I were lower than the Stanford-Binet scores by 7.21 points on the Full Scale, 8.46 points on the Verbal Scale, and 7.00 IQ points on the Performance Scale. For Group II, the corresponding WISC scores were lower than the Stanford-Binet by 5.00, 6.05, and 5.88 IQ points. The differences between these respective quantities, which represent the improvement attributable to practice, are 2.21 points for the Full Scale, 2.41 points for the Verbal Scale, and 1.12 IQ points for the Performance Scale. The fact that the Verbal Scale showed an improvement that is more than twice as great as that of the Performance Scale may be attributed to its greater similarity to the Stanford-Binet.

The writer has found only one other study, that of Kureth, Muhr, and Weisgerber,[12] which reported an investigation of the practice effect of these two scales. These investigators found no practice effect. The question must, therefore, remain open until more evidence is reported.

12. Sister Genevieve Kureth, Jean Muhr, and C. A. Weisgerber, Some data on the validity of the Wechsler intelligence Scale for Children. Child Develpm., 1952, 23, 281.

TABLE XVI

IMPROVEMENT IN PERFORMANCE UPON THE ADMINISTRATION OF THE SECOND OF TWO INTELLIGENCE SCALES

Group	Mean S-B IQ	WISC Full Scale Mean IQ	Diff. From S-B	WISC Verbal Scale Mean IQ	Diff. From S-B	WISC Perf. Scale Mean IQ	Diff. From S-B
Group I (WISC First)	113.42 (S.D.: 12.44)	106.21 (S.D.: 10.68)	7.21	104.96 (S.D.: 10.46)	8.46	106.42 (S.D.: 11.46)	7.00
Group II (S-B First)	114.88 (S.D.: 11.14)	109.88 (S.D.: 9.56)	5.00	108.83 (S.D.: 7.24)	6.05	109.00 (S.D.: 13.56)	5.88
Improvement			2.21		2.41		1.12

CHAPTER V

SUMMARY AND CONCLUSIONS

The Wechsler Intelligence Scale for Children and the Stanford-Binet, Form L, were administered in counterbalanced order to 48 normal white children (age 7 years 15 days to 8 years 11½ months) in regular grade at a representative public school in Detroit, Michigan. The data were analyzed by correlation and computation of the differences of means. A study was also made of individual discrepancies in test scores for seven-year-olds, eight-year-olds, and the combination of the two groups.

1. The coefficients of correlation of the Stanford-Binet and the three WISC quotients for the seven-year-olds are .759 for the Full Scale, .740 for the Verbal Scale, and .517 (Sign.: 2 per cent) for the Performance Scale. For the eight-year-olds, the corresponding coefficients were .638, .632, and .483 (Sign.: 5 per cent). The Standard Error for all of the above is .209. The corresponding coefficients for the two groups combined are .698, .710, and .529. The Standard Error for these is .146. All coefficients of correlation are significant at the 1 per cent level of confidence or better, except as otherwise indicated.

2. The mean Stanford-Binet scores are higher than those of the three WISC quotients in every group. The mean differences at the seven-year-old level are 6.37 IQ points for the Full Scale, 8.87 points for the Verbal Scale, and 5.66 (Sign.: 5 per cent) IQ points for the Performance Scale. The corresponding mean differences for the eight-year-olds are 5.83, 5.62, and 7.21; those for the combined groups are 6.11, 7.25 and 6.44 IQ points. All the differences are significant at the 1 per cent level of confidence or less, except the one indicated.

A similar comparison of mean differences in scores between the one-half of the subjects testing highest on the Stanford-Binet shows that the Stanford-Binet tested higher by 10.06 IQ points at the seven-year level, 11.88 points higher at the eight-year level, and 10.67 IQ points higher at the seven- and eight-year level. All are significant at the 0.1 per cent level of confidence, except that of the eight-year-olds, which is significant at the 1 per cent level. Corresponding mean differences for the one-half testing lowest on the Stanford-Binet were 1.00, 2.81, and 1.54 IQ points. These differences are not significant at the 20 per cent level of confidence, except that of the eight-year-old group.

3. Discrepancies in individual scores are found in 96 per cent of the cases. Seventy-three per cent of the subjects scored higher on the Stanford-Binet and 23 per cent had higher WISC Full Scale scores. Fifty-six per cent of

the subjects had Stanford-Binet scores that exceed the WISC scores by five IQ points or more, 35 per cent had Binet scores that were higher by ten or more points, and 19 per cent had Binet scores that were higher by fifteen points or more. In regard to higher WISC scores, 10 per cent of the whole group had differences of five or more IQ points and 4 per cent had WISC Full Scale scores that exceeded the Stanford-Binet scores by ten points or more. Individual score discrepancies ranged from one to twenty-five IQ points higher on the Stanford-Binet and from one to eleven points higher on the WISC Full Scale.

4. A regression equation computed for the conversion of Stanford-Binet scores into WISC scores was found to be: y equals .601x plus 39.4, in which y is the WISC Full Scale score and x is the Stanford-Binet score. The Standard Error is 7.35 IQ points, which is too large to allow one much confidence in the accuracy of the equation's predictive value. Prediction errors as large as 16 IQ points occurred through the use of the equation in cases where the actual discrepancy between individual scores was large.

5. When the WISC was administered first, the Stanford-Binet was higher than the Full Scale IQ by 7.21 points, higher than the Verbal Scale IQ by 8.46 points, and higher than the Performance Scale IQ by 7.00 points. When the Stanford-Binet was administered first, the corresponding WISC scores were lower by 5.00, 6.05, and 5.88 IQ points. The improvement, which may be attributed to practice, is

2.21 for the Full Scale, 2.41 for the Verbal Scale, and 1.12 IQ points for the Performance Scale.

Conclusions

1. The absence of a higher degree of correlation of the Stanford-Binet scores with the three quotients yielded by the Wechsler Intelligence Scale for Children is attributable to differences in the scales, rather than to fortuitous factors. The scales either measure different aspects of intelligence or measure the same factors in different proportions. These scales are, therefore, not interchangeable. On the other hand, the contributions of the WISC Full Scale and the Stanford-Binet are not sufficiently unique to justify the inclusion of both in a battery of tests used in routine clinical practice. The worker has the advantage of a choice between two instruments but also has the responsibility of selecting the one which, in his professional opinion, will best serve the clinical purpose.

2. With children of seven and eight years of age, the Stanford-Binet has a strong tendency to produce higher scores than the WISC at the upper levels of intelligence. The mean difference will be about ten IQ points for those children having Stanford-Binet scores of 114 points or higher. As the Stanford-Binet scores descend from the upper levels of intelligence toward the norm of IQ 100, the differences in mean scores become smaller. However, widely discrepant individual scores are likely to be found at any

intelligence level above IQ 90. It is impossible to predict, with a practical degree of accuracy, the probable individual score on one of these scales from the obtained score on the other.

3. The problems encountered in this study suggest the need for additional investigation to:

A. Determine if the conclusions reached in this study are valid for other samples, both those with similar characteristics and those in which the intelligence level, age, area of residence, training, and background are different.

B. Discover the casual factors, either in the instrument or the mental structure and personality of the subject, that could account for the discrepancy in scores.

C. Determine, through analytic studies, the comparative calibration of the two scales for subjects of different levels of intelligence as well as those of different training and background.

D. Provide, if possible, some objective criteria upon which to base the selection of one scale rather than the other in assessing the intellectual capacities of a specific individual.

BIBLIOGRAPHY

Frandsen, Arden N. and Higginson, Jay B. The Stanford-Binet and the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 136-138.

A presentation of the correlation and the means of the score yields by both instruments for 54 unselected, fourth-grade children. Special attention is given to the efficiency of each scale in forecasting school achievement.

Grove, William R. Mental Age scores for the Wechsler Intelligence Scale for Children. J. clin. Psychol., 1950, 6, 393-397.

A criticism of the abandonment of the Mental Age concept and presentation of a rather cumbersome method of estimating the MA for the Wechsler Intelligence Scale for Children. The author notes that errors may be produced in obtaining the IQ because of the gross age scaling on the test. An illustrative case is given.

Krugman, Judith I., Justman, Joseph, Wrightstone, J. Wayne, and Krugman, Morris. Pupil functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 475-483.

The most extensive and thorough study published to date. The results are based on the scores of 332 children attending representative New York City schools. An excellent study by competent investigators.

Kureth, Sister Genevieve. The validity of the subtests of the Wechsler Intelligence Scale for Children with five- and six-year-olds. Unpublished Master's Thesis, University of Detroit, 1953.

A study of the validity of the Wechsler Intelligence Scale for Children by the correlation of the subtests. The basic data were used in the published study by Kureth, Martin, and Weisgerber, listed below.

Kureth, Sister Genevieve, Muhr, Jean P. and Weisgerber, Charles A. Some data on the validity of the Wechsler Intelligence Scale for Children. Child Develpm., 1952, 23, 281-287.

Using the combined basic data of the first two investigators, the senior author elaborated on the studies. The larger sample and the extended statistical treatment lends greater weight to the conclusions. The emphasis is on the size and frequency of the discrepancies in scores yielded by the Stanford-Binet and the Wechsler Intelligence Scale for Children.

Muhr, Jean P. Validity of the Wechsler Intelligence Scale for Children at the five and six year level. Unpublished Master's Thesis, University of Detroit, 1952.

A validation study using the Stanford-Binet as the criterion. The sample consisted of forty-two normal, institutional children. The basic data were used in the published study cited above.

Pastovic, John J. and Guthrie, George M. Some evidence on the validity of the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 385-386.

The basic data from three unpublished master's theses and one published study, together with the scores of two groups tested by the junior author are presented and interpreted. The Stanford-Binet and the Arthur Performance Scale are used as criteria.

Rapaport, I. A comparison of performance on the Wechsler Intelligence Scale for Children and the Revised Stanford-Binet Scale. In Pastovic, John J. and Guthrie, George M. Some evidence on the validity of the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 385-386.

One of the unpublished master's theses reported in Pastovic and Guthrie, above.

Seashore, Harold G., Wesman, Alexander, and Doppelt, Jerome. The standardization of the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1950, 14, 99-110.

Essentially an expansion of the technical section of the WISC manual, dealing with the standardization sample, the intercorrelation of the Verbal and Performance Scales, and the performance of certain particular groups on the test.

Seashore, Harold G. Differences between verbal and performance IQ's on the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 62-67.

Using the standardization data as a basis, the author presents an estimate of the frequency and size of the discrepancies that may be expected between the scores yielded by the Verbal and Performance Scales of the Wechsler Intelligence Scale for Children. He offers his opinions on the significance that is to be attached to such differences.

Sloan, William, and Schneider, Bernard. A study of the Wechsler Intelligence Scale for Children with mental defectives. Amer. J. ment. Def., 1951, 55, 573-575.

A study to determine the relationships between the Wechsler Intelligence Scale for Children, the Arthur Performance, and the Stanford-Binet in a group of 40 high grade mental defectives. The Wechsler norms at the lower end of the scale are criticized.

Stacey, Chalmers L. and Levin, Janice. Correlation analysis of scores of subnormal subjects on the Stanford-Binet and Wechsler Intelligence Scale for Children. Amer. J. ment. Def., 1951, 55, 590-597.

Results are presented of the performance of two groups consisting of 44 morons and 26 borderline defectives in a state school. The four quotients are correlated and the values of the mean scores are presented for each group separately and both groups together. Many significant differences are reported.

Terman, Lewis M. and Merrill, Maud A. Measuring Intelligence. Boston: Houghton Mifflin Company, 1937.

The guide to the administration and scoring of the Stanford-Binet, together with a report on its standardization.

Wechsler, David. The Measurement of Adult Intelligence. Third edition. Baltimore: The Williams and Wilkins Company, 1944.

This manual gives a reasonably comprehensive statement of Wechsler's concept of the nature of intelligence and the theories underlying his tests.

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

The manual of instructions for the administration and scoring of the Wechsler Intelligence Scale for Children. The technical section is sketchy and incomplete.

Wechsler, David. Equivalent Test and Mental Ages for the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 381-384.

Three methods are presented for determining the Mental Age equivalent of test scores.

Weider, Arthur, Noller, Paul A., and Schramm, Theodore A. The Wechsler Intelligence Scale for Children and the Revised Stanford-Binet. J. consult. Psychol., 1951, 15, 330-333.

On the basis of the scores of 106 normal Louisville children, the authors developed, by the use of a regression formula, a table of equivalent scores for the two tests. The table cannot be used with a high degree of confidence.