

LE 3 B7  
1951 A8  
P58 C6  
cap. 1

A COMPARISON OF THE STANFORD-BINET (1937 REVISION,  
FORM L) AND WECHSLER INTELLIGENCE SCALE FOR CHILDREN  
AT DIFFERENT AGE AND INTELLECTUAL LEVELS

by

JOAN ANNE POWELL

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF ARTS  
in the Department  
of  
PHILOSOPHY AND PSYCHOLOGY

We accept this thesis as conforming to the  
standard required from candidates for the  
degree of MASTER OF ARTS

Members of the Department of  
Philosophy and Psychology.

THE UNIVERSITY OF BRITISH COLUMBIA

August, 1951

A COMPARISON OF THE STANFORD-BINET (1937 REVISION,  
FORM L) AND WECHSLER INTELLIGENCE SCALE FOR CHILDREN  
AT DIFFERENT AGE AND INTELLECTUAL LEVELS

Abstract

This study was designed to investigate the relationship between the intelligence quotients yielded by two widely used individual tests of intelligence for children, namely, the Stanford-Binet, Form L, (1937 Revision) and the Wechsler Intelligence Scale for Children (WISC). Until recently, the Binet has been used almost exclusively to ascertain the intelligence of the school-age child but, with the publication of the WISC in 1949, there has been an increasing trend toward using the tests either interchangeably or in conjunction with one another. In view of this development, an attempt to discover the relationship between the two scales would seem to be of much practical value.

Although the two scales agree in assuming a "g" factor of intelligence, they differ as to the nature of their content and construction. The Stanford-Binet does not include any test items designated as measuring a particular skill, whereas the WISC is composed of twelve subtests, each supposed to tap a specific ability, and it yields a separate verbal and performance intelligence quotient. The two scales also differ in the manner of computing an intelligence quotient; the Binet scale depends upon a Mental Age concept of intelligence, whereas the Wechsler Intelligence Scale for Children is a point scale.

This study attempted to ascertain to what extent the Stanford-Binet correlates with each of the WISC scales (Verbal, Performance and Full Scale) at three different age levels and three levels of intelligence, i.e., with subjects of Superior, Average and Retarded intelligence. It also attempted to find out what differences, if any, might occur between the Mean intelligence quotients yielded by the two tests in the above age and intellectual categories, and what direction these differences might take.

Wechsler has objected to the Stanford-Binet deviations, which vary in size at difference age levels. At 6 years, the Binet standard deviation is unusually small, and at 12 years of age it is unusually large: the WISC standard deviations are the same size at

each age level. It was hypothesized, therefore, that at the extremes of the intelligence distribution at ages 6 and 12 years, there should be differences between the Mean intelligence quotients yielded by the two tests in the direction of the size of the Binet standard deviations at these two age levels - a smaller Mean Binet than Mean WISC intelligence quotient at age 6 years, with a higher Mean intelligence quotient on the Binet at age 12 years. Subjects of these two ages, 6 and 12 years, were included in the experimental group in order to test this hypothesis, while the use of subjects of superior and retarded intelligence insured that extreme scores would occur. The sample of subjects of average intelligence, plus a group of 9-year-olds, were included for control and comparison in testing this hypothesis but also for their own research value.

The sample was composed of 85 subjects - ten children in each age category of the Superior and Average intelligence groups; and in the Defective group, fourteen 12-year-olds, nine 9-year-olds, and four 6-year-olds.

The positive correlations which occurred may be summarized as follows :

1. In the 9-year-old Superior group, the Stanford-Binet IQ correlated significantly with -
  - (a) the WISC Verbal IQ at the 1% level of confidence;
  - (b) the WISC Performance Scale IQ at the 5% level of confidence;
  - (c) the WISC Full Scale IQ at the 1% level of confidence.
2. In the 9-year-old Average group, the Stanford-Binet IQ correlated significantly with -
  - (a) the WISC Verbal Scale IQ at the 1% level of confidence;
  - (b) the WISC Full Scale at the 5% level of confidence.

Significant differences between the Mean IQs of the two tests may be summarized as follows :

1. In the group of Superior 9-year-olds, the Stanford-Binet IQs were significantly higher (at the 1% level of confidence) than the WISC Verbal, Performance, and Full Scale IQs.

2. In the group of Superior 12-year-olds, the Stanford-Binet IQs were significantly higher at the 1% level of confidence for the WISC Full and Verbal Scale IQs.
3. In the group of Average 12-year-olds, the Stanford-Binet is significantly higher at the 5% level of confidence than the WISC Verbal IQ.

The major conclusions of this study are :

1. The obtained results are in essential agreement with the studies comparing the Wechsler adult scale and the Stanford-Binet.
2. There seems to be a consistent tendency in this study and others reviewed previously for lower correlations between the Stanford-Binet and WISC Performance Scale, than between the Stanford-Binet and WISC Verbal and Performance Scales.
3. There seems to be no support for the hypothesis that the difference between the Mean Stanford-Binet and the Mean WISC IQs at the Superior level will differ in direction according to the size of the Binet standard deviation at the age level in question.
4. The WISC appears to be an unsatisfactory test for measuring the markedly retarded children. Both in terms of construction, and interest value to subjects, the Stanford-Binet seems to be a better scale for the measurement of the lower levels of intelligence.
5. Keeping in mind the limited sample upon which this study was based, the two scales do not seem to be interchangeable. The practical import of this conclusion is that clinicians, social workers, psychiatrists, school teachers, and so on, should be fully aware that the child given both the tests may well yield widely different IQs on the respective tests.

Suggestions for future research have been included.

## C O N T E N T S

CHAPTER		PAGE
I	INTRODUCTION	1
	Statement of the Problem	1
	Historical Background	1
	Review of the Literature	3
II	PROCEDURE	13
	Selection of Subjects	13
	Administration of the Intelligence Scales	16
III	TREATMENT OF DATA AND RESULTS	22
	Treatment of Data	22
	Results : (a) Obtained Correlations	22
	between Scales	
	(b) Comparison of Mean IQs	28
	(c) The Performance of the	41
	Defective Group	
IV	DISCUSSION OF RESULTS	47
	Theoretical Expectations and Obtained	47
	Results	
	Obtained Results compared with the	56
	Literature	
	Consistent Trends emerging from the Data	58
	Suggestions for Future Research	58
V	SUMMARY	60
	BIBLIOGRAPHY	64

# T A B L E S

TABLE		PAGE
I	Original Plan for selection of Sample	14
II	Obtained Sample represented as to Age and Intelligence Level Groups	17
III	Correlations between Stanford-Binet and WISC Verbal, Performance, and Full Scales	24
IV	Correlations between Stanford-Binet and WISC Verbal, Performance, and Full Scales	26
V	Mean WISC Verbal Scale IQs compared with Mean Stanford-Binet IQs	29
VI	Mean WISC Performance Scale IQs compared with Mean Stanford-Binet IQs	30
VII	Mean WISC Full Scale IQs compared with Mean Stanford-Binet IQs	31
VIII	Mean WISC Verbal Scale IQs compared with Mean Stanford-Binet IQs	36
IX	Mean WISC Performance Scale IQs compared with Mean Stanford-Binet IQs	37

TABLE		PAGE
X	Mean WISC Full Scale IQs compared with Mean Stanford-Binet IQs	38
XI	Performance of 12-year-old Mental Defectives on WISC Verbal, Performance, and Full Scales, and on Stanford-Binet	44
XII	Performance of 9-year-old Mental Defectives on WISC Verbal, Performance, and Full Scales, and on Stanford-Binet	45
XIII	Performance of 6-year-old Mental Defectives on WISC Verbal, Performance, and Full Scales, and on Stanford-Binet	46

## ACKNOWLEDGEMENTS

The author is most grateful to her Faculty adviser, Mrs. D. T. Kenny, for her continued encouragement and many helpful suggestions and criticisms.

The author is indebted to Dr. R. F. Sharpe, Acting Assistant Superintendent and Inspector of Secondary Schools, for permission to test in the Vancouver Elementary Schools. She is also grateful to Dr. L. L. Sauriol, Superintendent of The Woodlands School, New Westminster, B.C., and to Dr. T. D. Barber, Superintendent of The Rainier State School, Buckley, Washington, U.S.A., for making available the retarded subjects used in this study.



## CHAPTER I

### INTRODUCTION

#### Statement of the Problem

The purpose of this study is to investigate the relationship between the intelligence quotients yielded by two widely used standard intelligence scales, namely the Stanford-Binet Form L (1937 Revision), and the Wechsler Intelligence Scale for Children (WISC). Until recently, aside from group intelligence tests, the Binet scale has been used almost exclusively to ascertain the intelligence of the school-age child. With the emergence of the WISC in 1949, there has been an increasing shift toward the use of this scale either in conjunction with, or in place of, the Binet scale. In view of this trend, a comparison of the two scales, such as is proposed in this study, is of much practical value. That is, if the tests are to be used interchangeably, some attempt to demonstrate the relationship between the two scales is urgently needed.

#### Historical Background

Although the authors of both of these scales follow

Spearman in assuming a global factor of intelligence, the two scales are quite different in the nature of their construction. No attempt is made in the Binet scale to group the items in terms of the kinds of skills involved, whereas the Wechsler Intelligence Scale for Children is composed of twelve subtests, each of which involves a particular skill, and the scale yields a separate Verbal and Performance intelligence quotient. In this respect, the Wechsler scale is sometimes thought to lean toward Thurstone's approach to intelligence in terms of special abilities. Wechsler himself, however, is careful to point out that he is primarily interested in arriving at a general measure of intelligence and that the subtests are not "...a series of tests that measure primary abilities," (p.5, 20).

Besides differing as to the type of material and its arrangement, the two tests show sharp differences in the method of arriving at the intelligence quotient. The Binet scale depends on a Mental Age concept of intelligence, whereas the Wechsler scale falls into the category of a point scale. Wechsler has voiced strong objections to the Stanford-Binet  $\frac{MA}{CA} = IQ$  formula for several reasons, to be discussed later in this paper. Both the adult, and the more recent Wechsler Intelligence Scale for Children, yield an intelligence quotient resulting from a comparison of the subject's performance with that of others of his own age, rather than the equating of the performance with a certain Mental Age level. Wechsler justifies this move on several grounds. Firstly, he points out that a Mental Age of 7

years is not at all the same intellectual capacity in a 10-year old or a 5-year old as it is in a 7-year old, as is implied by the construction of the Binet scale. Secondly, he questions the assumption, innate in the Stanford-Binet, that intellectual growth stops at 16 years. His method of computing the IQ makes any such assumption unnecessary; it also makes allowances for the slowed-up intellectual development of puberty and the early teens, and for the decrease in intellectual ability which occurs with age. Whether or not this method, which is more lenient to older subjects, is better depends upon the use to which the score will be put. Thirdly, Wechsler's chief objection to the Stanford-Binet is to the size of the standard deviations at different age-levels. The greatest deviations occur at age 12 years ( $\sigma$  LIQ = 20.00), and at 6 years ( $\sigma$  LIQ = 12.50). He points out (p.26, 19) that a child 2 sigma away from the Mean at age 6 years will have an IQ of 75, whereas a child similarly placed at 12 years will have an IQ of 60. Hence the scale yields a rather unreliable estimate of the subject's intelligence.

### Review of the Literature

Early attempts to investigate the relationship between intelligence quotients derived from Mental Ages and those derived from point scales have been based on comparisons of the Stanford-Binet with the Wechsler-Bellevue Adult Intelligence Scales. Since the Wechsler Intelligence Scale for Children, used in this study, has grown out of

and closely parallels the Adult Scale, a review of the above mentioned studies seems in order at this point.

A number of these studies were done with adolescents since it is in this age range that both scales are widely used. Sartain (13) compared IQs obtained on the Stanford-Binet with those obtained on the Wechsler-Bellevue by fifty adolescent college freshmen. He found that the Stanford-Binet correlated with the Full Scale, .774; with the Verbal Scale, .802; and with the Performance Scale, .510. He estimates that the Stanford-Binet yields an IQ approximately 5 points higher than does the Wechsler-Bellevue; The Mean Binet IQ was 129.44, and the Mean Wechsler IQ was 117.44 with a critical ratio of 5.55. No definite conclusions are stated by the author and the results are ambiguous because one cannot be sure which of the variables, (a) adolescence, (b) superior intelligence, or, (c) both, accounts for the results.

Goldfarb (6) used a sample of sixty adolescents in foster homes, with Mean Age equal to 14.6 years. The Mean Verbal IQ was highest for 62% of the subjects. The Full Scale Wechsler correlated with the Stanford-Binet to produce a coefficient of .86; the Verbal, a coefficient of .80; and the Performance, a coefficient of .67. The author claims that bright subjects test higher on the Revised Stanford-Binet than on the Wechsler, while the advantages are reversed for the dull subjects. The author does not qualify what he means by "bright" and "dull". He further states that younger children get higher

Stanford-Binet IQs and older subjects higher Wechsler IQs, but again does not specify as to what these categories encompass. He also maintains that the Wechsler does not discriminate very well amongst groups of superior adolescents. The fact that these subjects were in foster homes does not appear to have produced results different from those that might have been expected from a comparable sample not having this characteristic.

Guertin (7) refers to an unpublished study of his own with a sample of feebleminded subjects with IQs between 51 and 75, and between the ages of 15 and 22 years. He found that the Wechsler-Bellevue EQ for these subjects was 5.33 points above the Binet IQ, although he does not state whether or not this difference was significant.

From these studies it appears that adolescents of superior intelligence achieve higher IQs on the Stanford-Binet, and that dull adolescents achieve higher IQs on the Wechsler. Younger adolescents seem to get higher Binet than Wechsler IQs, and older adolescents seem to get higher Wechsler IQs.

Kutash (10) used a sample composed of fifty adult mental defectives. He found that (a) the Wechsler-Bellevue yielded higher IQs in 84% of the cases, (b) the Mean Wechsler IQ was 11 points higher than the Mean Stanford-Binet IQ, the difference significant at the 1% level of confidence, (c) the size of the difference in IQs varies directly with the chronological age of the subject, and (d) the two

scales correlated,  $r = .77$ . He concludes that differences in the IQs obtained on the two scales are due to differences in norms and principles of standardization. The standardization and norms of the Wechsler take into account the normal deterioration with age which the Binet does not do. Rabin (12) and Guertin (7) point out that the results of this study may not be valid owing to the lack of homogeneity as to age in the group selected.

Halpern (9) conducted a study using one-hundred-and-thirty-three patients at a mental hygiene clinic. The subjects were all classified as having dull normal intelligence, and were divided into four age groups. For ages 15 to 34 years, the tests yielded similar measures, although Halpern prefers the Wechsler for several technical reasons. It was found that from ages 10 to 14 years, higher IQs are consistently yielded by the Stanford-Binet. Halpern believes that for subjects who are 13 years and younger, and for 14-year-olds who are of low intelligence, the Binet is the preferable instrument. He suggests that the Binet IQs are too high and that the Wechsler IQs are too low. When he took a group of subjects and divided them according to intelligence, the highest correlations were at the extremes. In spite of this fact, the greatest differences occurred with the superior subjects, and the least with the retarded subjects. He would postulate that "...both scales tap the defective's limited capacity equally well, but that the range of the superior subject's is reached equally by all tools," (p.210, 9). Because the Wechsler norms

take the deterioration of intellectual ability with age into account, he says that differences between the IQs in those over 34 years are generally accountable for in terms of differences in test construction. He believes that the introduction of the non-verbal tasks in the Wechsler-Bellevue makes for further test differences.

Balinsky, Israel and Wechsler (2) tested the relative effectiveness of the Wechsler-Bellevue and the Stanford-Binet in diagnosing mental deficiency. The criterion was psychiatric diagnosis. The best predictions were made in this order : the Wechsler Full Scale, the Wechsler Verbal Scale, the Stanford-Binet, and, last, the Wechsler Performance Scale. Using bi-serrial r's, the authors computed the forecasting ability of the Wechsler to be 40%, and that of the Binet to be 5%. The sample was of patients in the Bellevue Hospital. Since the Wechsler-Bellevue was validated clinically in the Bellevue Hospital (p.127, 19), and, presumably, according to the standards of that institution, and since this study was carried out in the same institution, it is not surprising that Wechsler scores of a subsequent sample should agree more closely with those standards than scores on a test validated on other criteria.

In summary, it would appear that retarded subjects attain higher IQs on the Wechsler-Bellevue than on the Stanford-Binet. This could result from the fact that, as Wechsler points out (p.157, 19), the feeble-minded generally do better on the Performance than on the Verbal Scale. Differences would be expected at this level of intelligence,

since, of all the three Wechsler scales, the Performance Scale correlates lowest with the Binet, and a higher IQ would be expected because a better Performance index would naturally tend to raise the Full Scale IQ.

Benton, Weider and Bleauvelt (3) studied a sample of sixty subjects described as "mental cases" who were cooperative. The subjects ranged in age from 16 to 59 years, with a Mean Age of 35 years. The correlations were all high, although, as usual, the Performance Scale showed the lowest agreement. The coefficients were as follows : Full Scale, .93; Verbal Scale, .92; and Performance Scale, .73. Although the Mean scores are similar, the distributions differ greatly, the standard deviation of the Stanford-Binet far exceeding that of the Wechsler. The authors feel that these disparities do not indicate a real difference in what the tests measure because widely diverse scores on the Binet and Wechsler may indicate similar positions in relation to the Means. This becomes obvious when the high correlations are remembered. The difference is, again, due to the Wechsler norms; the authors express the opinion that the scores would be comparable if converted into percentile ranks or standard scores. This study must be regarded as inconclusive since the nature of the sample is so ill-defined.

Mitchell (11) used a sample of two-hundred-and-sixty-eight subjects, mostly delinquents and chronic alcoholics without psychosis, and cooperative psychotics. He found that all three parts of the Wechsler correlated significantly with the Binet; twenty-one



of his subjects showed a difference of more than 20 IQ points between their scores on the two scales; of these, sixteen showed a higher IQ on the Wechsler. The ages of these subjects ( $M = 48.2$  years) was nearly double that of the five subjects who scored 20 points higher on the Stanford-Binet. Since the sample is heterogeneous, any characteristics which might have accrued to one clinical group may have been obscured by the opposite tendencies in another group; no justification is given by the author for the sample selected.

On the basis of these studies, it appears that superior subjects do better on the Stanford-Binet than on the Wechsler, and that retarded subjects do better on the Wechsler than on the Binet. Up to the early teens, higher scores are found on the Binet, and, after this age level, higher scores are found on the Wechsler. In all studies, the Wechsler Performance Scale correlates lower with the Stanford-Binet than do the Verbal or Full Scales. With the exception of those studies involving adolescents, all these comparisons are at a disadvantage. They are attempting to compare IQs on two tests, one of which was standardized on adults and the other on children.

Literature on the Wechsler Intelligence Scale for Children has been scant to date. Of the four articles published on it, only two bear on this investigation. The first, by Seashore, Wesman and Doppelt (14), is concerned chiefly with describing the standardization of the WISC, a matter which is not of great interest here. The authors do point out, however, that it is unlikely that there will be very

extreme scores because the range of IQs on a point scale such as the WISC is quite arbitrary. If the standard deviation had been set at 20 IQ points rather than the 15 points actually used, extreme scores might be expected.

The only other study that is related to this investigation was that done by Frandsen and Higginson (5), who compared IQs on the WISC with those on the Stanford-Binet. The scales were also tested for validity in predicting school success. The sample included fifty-four Fourth Grade children between the ages of 9-years-and-1-month and 10-years-and-3-months; the subjects were of average ability and average achievement. The Full Scale WISC and Stanford-Binet correlated to yield a coefficient,  $r = .80$ , and the Verbal and Performance Scales yielded coefficients of .71 and .76 respectively. The validity of the Binet for predicting school success as measured by the Standard Achievement Test was found to be .63, and that of the WISC to be .76. Only in the Language Achievement subtest did the Stanford-Binet predict better than the WISC.

Unlike the studies reviewed comparing the Stanford-Binet and Wechsler-Bellevue, this study by Frandsen and Higginson answers little concerning test differences in any except the average intellectual group. As indicated by the literature on Wechsler-Bellevue and Stanford-Binet studies, large differences do not occur in the groups of average subjects, but do occur with the superior subjects and with the retarded ones, and with the older subjects and with the younger

ones. On the basis of these same studies, for instance, one might expect that retarded subjects would achieve higher WISC than Binet IQs, and that the superior subjects would have lower WISC than Binet IQs. In the present investigation, an attempt will be made to compare the performance of subjects of superior and defective intelligence, as well as those of average intelligence.

In the selection of age levels to be used, this study has an advantage over the studies comparing the Wechsler-Bellevue with the Stanford-Binet, in that both of the present tests have been standardized on children.

The WISC is a point scale with a set standard deviation, while the size of the Stanford-Binet standard deviation varies, being unusually small at age 6 years and unusually large at age 12 years. As was discussed earlier, Wechsler argues that this characteristic of the Binet produces unreliable results, since subjects at the top and bottom parts of the normal curve show a different IQ at different age levels. At the extremes of the normal curve, i.e., with the retarded and the superior subjects at these ages, therefore, there should be a Mean difference between the IQs produced by the two scales. Since the WISC has a set standard deviation, the same at each age level, the Mean Differences should be in the direction corresponding to the size of the Binet deviation at that level. These ages, 6 and 12 years, will be included in the sample, therefore, in order to test this hypothesis. The inclusion of samples of superior and retarded

children will insure that extreme scores (i.e., at either end of the normal curve) will occur. This will give information omitted by Frandsen and Higginson, for, although they report high correlations, they do not say whether the scores are comparable as to size. It will also give information as to whether the high agreement they report also holds for other age levels.

Besides the 6 and 12-year-olds, the sample will include a group of 9-year-olds, since the Binet standard deviation at age 9 is not only near the average for that test (it is equal to 16.4 IQ points and the average is 16 points) but it also approximates the WISC deviation of 15 IQ points. The 9-year-olds, plus a sampling of average children, will be included for their own research value, but mainly for purposes of control and comparison.

In brief, this investigation is planned to find out to what extent the three indices of intelligence yielded by the WISC on the Full, Verbal and Performance Scales, correlate with those yielded by the Stanford-Binet at ages 6, 9 and 12 years, using superior, average and defective children. It will endeavor to ascertain whether the Mean scores in these groups show real differences in average IQs yielded by the two tests, and, if so, in what direction these differences do occur.

## CHAPTER II

### PROCEDURE

In order to determine the extent to which intelligence quotients by the Binet scale (1937, Revised edition) and the Wechsler Intelligence Scale for Children are comparable, the following method of investigation was employed.

#### Selection of Subjects

The study was planned so that the sample should include thirty children of superior intelligence, thirty of average intelligence, and thirty retarded children. Each group of thirty would include ten children aged 12 years, ten aged 9 years, and ten aged 6 years. Thus, the sample should include thirty children at each age level. In this way, varying intelligence and varying age levels were represented in the sample used in this study. The total sample should equal ninety. The planned selection is represented in Table I, (p. 14).

The 9 and 12-year-old Superior subjects were obtained from various elementary schools throughout Vancouver. They were

TABLE I

## ORIGINAL PLAN FOR SELECTION OF SAMPLE

N = 90

Age	Intellectual Level			
	Superior	Average	Defective	Total
6 years	N = 10	N = 10	N = 10	= 30
9 years	N = 10	N = 10	N = 10	= 30
12 years	N = 10	N = 10	N = 10	= 30
Total	= 30	= 30	= 30	= 90

selected on the basis of scores on either the Detroit First Grade Examination, or Otis Score, or scores on the National Intelligence Scale, or on all of them, depending upon what tests the child had taken. The sample does not include any subject whose academic standing is at variance with the level of ability indicated by the intelligence tests, nor does it include any subjects for whom the results of different tests were equivocal. Some of the 6-year-old Superior subjects were selected on the basis of teachers' reports: these evaluations proved unreliable, however, and additional subjects had to be tested in order to obtain the required number of subjects at this level.

The subjects in the Average category were selected on much the same basis as the Superior group, that is, scores on previous intelligence tests, grades and marks and class standing. No difficulty was encountered in obtaining the required subjects in this category.

The sample of feebleminded subjects was obtained from the Woodlands School, New Westminster, B.C., and the Rainier State School, Buckley, Washington, U.S.A. Besides fulfilling the age requirements, these children had to be defectives classified as "familial", and an attempt was made to use only the brighter children within the "moron" or "borderline" classification. However, several problems presented themselves in the work at this level. In the first place, neither of the two schools had a sufficient number of children at these age levels who were classified as "familial", and, in the second place, of those obtainable few were sufficiently

intelligent to meet the requirements of the study. These requirements were set up because Wechsler included in the retarded sample of his standardization group "...cases of the required ages who were rated as having IQs under 70 and not below 50," (p.8, 20). The third and largest problem occurred with the 6-year-old Retarded group. Except in unusual instances, such as cases of physical damage and mongolism, children are not admitted to the Woodlands School until they are 6 years of age. Few 6-year-olds were obtainable, therefore, and those who were showed very low intelligence. The reason for the great difficulty in obtaining subjects at this level may be in the ages selected. In other words, early committal to an institution may, generally speaking, be confined to cases of physical disability, mongolism and very great retardation. For these reasons, the sample of Retarded children has been curtailed, especially in the 6-year-old group. The obtained sample used in this study is indicated in Table II, (p. 17).

#### Administration of the Intelligence Scales

All the testing was done by the investigator who had been trained in the administration and scoring of both of the scales used in this study. Every effort was made to follow the standard procedures recommended in the test manuals. The records were scored first by the examiner and then the scoring was checked by a trained worker in the field.



TABLE II

OBTAINED SAMPLE REPRESENTED AS TO AGE  
AND INTELLIGENCE LEVEL GROUPS

N = 85

Age	Intelligence Level			
	Superior	Average	Defective	Total
6 years	N = 10	N = 10	N = 4	= 24
9 years	N = 10	N = 10	N = 9	= 29
12 years	N = 10	N = 10	N = 12	= 32
Total	= 30	= 30	= 25	= 85

Half the subjects in each age-intellectual group were given the Stanford-Binet first, and the other half were given the WISC first, in order to minimize and even out the effects of practice. The second test was given within a week, but never immediately after the first in order to avoid fatiguing the subject. The two tests were given as nearly as possible under the same conditions, and care was taken to prevent the testing running through any preferred activity of the child, e.g., recess or a school concert, or when the children were in a state of excitement, for instance, before Sports' Day.

In general, a high level of cooperation was obtained from the subjects. The school children, especially, participated most eagerly. In the case of the 6-year-olds, the examiner was introduced to the class and it was then explained that some of the children were to "play some games" with her. In this way, the testing became a most desirable and pleasurable thing to these youngsters.

Initially, these children were asked not to tell their friends the contents of the tests. However, this seemed to impart to the whole affair the atmosphere of a "secret" which the youngsters found most difficult to keep, so that after a while this practice was dropped. On the whole, whether or not the subjects were asked not to tell, very little evidence was shown of subjects knowing the test material beforehand.

Some of the subjects were asked to name the test they

preferred, but, in general, there did not seem to be a preference for either test.

More difficulty was encountered in establishing rapport with the feebleminded children, especially the very young ones. These children were usually more cooperative, however, when they were brought a second time after the first attempted interview had been abandoned. Occasionally, it seemed better to allow the attendant to remain in the room throughout the testing if the child were very shy. In these cases, the attendants were asked not to help the child in any way, and this request was honored without exception.

It was the impression of the investigator that, in the main, the WISC was much more difficult for the Retarded child than was the Binet. Such a distinction can easily be made simply in terms of establishing rapport and engaging the subject's attention. The little toys of the Stanford-Binet are immediately attractive to subjects of low mental age, but the WISC equipment does not have a vestige of this attractiveness. The statement - "These pieces, if put together correctly, will make a boy. Go ahead and put them together." - was, apparently, meaningless to most of the Retarded subjects. Obviously, if the subject fails to comprehend the nature of the task, he can not successfully complete it. Differences between the two tests in the practical value of a score at this level are illustrated very well in the Picture Completion Test. Instructions to tell what was missing only occasionally evinced a suitable response

from the subject, but most of these children could and did spontaneously name the pictures. This ability would add to the subject's score on the Binet, but, so far as a WISC score is concerned, it makes no difference whether the subject did name the pictures or merely sat in dull silence. For these reasons, the Stanford-Binet seems more suitable for use with Retarded children, and also because it does yield IQs at a lower level than does the WISC, although it did fail with most of the Retarded 6-year-olds and with some of the 9-year-olds in this study.

In administering the WISC to subjects of 8 years and older, not suspected mental defectives, the examiner is permitted to omit the first few items in each subtest and to start at a designated point of difficulty along the scale. If the subject does not achieve a certain number of consecutive correct responses, however, the examiner must work back from the designated point until the required number of consecutive correct responses has been given. While this technique is at times undoubtedly time-saving in testing older and brighter subjects, it is also somewhat awkward, especially if it is constantly necessary to move backwards to easier items. Also, some subjects notice the marked change in the level of difficulty of the items, and, in this way, become aware of having failed the more difficult items. Finally, the administration of the easier items in the Performance scale seems excessively complicated; for instance, different directions and examples are given with each of the first

four items in the Block Design and Picture Arrangement tests.

## CHAPTER III

### TREATMENT OF DATA AND RESULTS

#### Treatment of Data

In order to find out to what degree the three indices of intelligence, yielded by the WISC on the Verbal, Performance, and Full Scales, correlate with those yielded by the Stanford-Binet, at ages 6, 9 and 12 years, using Superior and Average subjects, Pearson Product Moment correlations were computed in each age-intellectual group between each Wechsler scale and the Stanford-Binet. An attempt is made to ascertain the nature of IQ differences yielded by the two tests by computing Mean Differences between the Stanford-Binet and Verbal, Performance, and Full Scales of the WISC in each age-intellectual group, and then using the t-test to test for significance of the difference.

#### Results

(a) Obtained correlations between scales :

##### Superior Subjects

The obtained coefficients of correlations between these two measuring instruments based on the Superior subjects are shown in

Table III, (p. 24).

With the 6-year-old Superior subjects, the Verbal Scale correlation with the Binet was the lowest in this age-intellectual group, and the third lowest in the entire Superior category; the correlation was not significant. The Performance Scale correlation coefficient is the highest in this age-intellectual group, but is not significant. The Full Scale correlation is slightly smaller than that of the Performance Scale; it is the median value for this entire intellectual category. In the 9-year-old Superior group, the Verbal Scale correlation is significant at the 1% level of confidence, and is the highest in the entire study. While the Performance Scale coefficient is the lowest in this age-intellectual group, it is significant at the 5% level of confidence. The Full Scale correlation is the second highest in the whole intellectual category, and is significant at the 1% level of confidence. The Verbal Scale correlation is the highest in the 12-year-old Superior group, and the correlation next to, but lower than, the median value for the entire intellectual category; the value, however, is not significant. The Superior 12-year-old Performance Scale correlation is the lowest in the entire study. The Full Scale correlation is the second lowest in the entire intellectual category; the coefficient is not significant.

#### Average Subjects

The correlations based on Average subjects are found

TABLE III

CORRELATIONS BETWEEN STANFORD-BINET  
AND WISC VERBAL, PERFORMANCE, AND  
FULL SCALES

Superior Subjects

Age	Verbal Scale	Perform- ance Scale	Full Scale
6 years	.34	.53	.52
9 years	.93	.67	.91
12 years	.50	.01	.33

$$.05 = .602$$

$$.01 = .735$$



in Table IV, (p.26).

In the 6-year-old Average group, the Verbal Scale correlation is not significant and is the lowest in the group, and the third lowest in the average intellectual group. The Performance Scale shows the highest correlation in this age-intellectual group, although the coefficient is not significant. The Full Scale correlation is slightly lower than the Performance Scale correlation; it is the median value for the entire Average intelligence group. The Verbal Scale correlation at the average 9-year-old level is the highest for the entire Average group, and is significant at the 1% level of confidence. The Performance Scale coefficient is the lowest in this age-intellectual group, and the second lowest in the entire Average category; it is not significant. The Full Scale correlation is second highest in this age-intellectual group and in the entire Average category; it is significant at the 5% level of confidence. In the 12-year-old Average group, the Verbal Scale shows the highest correlation; it is the same as the Average 6-year-old Full Scale coefficient, and is also the median value for the Average group; the value is not significant. The Performance Scale correlation is the lowest for this age-intellectual group and for this entire intellectual category. The Full Scale correlation for this age-intellectual group is the medium one and is not significant.

Summarizing the results of Tables III and IV, it is immediately apparent that in both intellectual groups, the rank order of the correlations is the same at the same ages. The highest, second

TABLE IV

CORRELATIONS BETWEEN STANFORD-BINET  
AND WISC VERBAL, PERFORMANCE, AND  
FULL SCALES

Average Subjects

Age	Verbal Scale	Perform- ance Scale	Full Scale
6 years	.31	.52	.50
9 years	.77	.28	.73 <sup>+</sup>
12 years	.50	.24	.46

$$.05 = .602$$

$$.01 = .735$$

<sup>+</sup>All quantities are rounded off to second decimal place. When worked to third decimal place,  $r = .727$ .

highest, median, third lowest, and lowest correlations for each intellectual group occur with the same WISC scales (i.e., Verbal, Performance, or Full Scales) at the same age levels. In each age-intellectual group, the Full Scale correlation is the middle value, and in both 6-year groups the rank order of the Verbal and Performance Scales is the reverse of what it is in both of the 9-year and both of the 12-year groups.

The rank order of the Verbal correlations is the same in both intelligence groups, with the 9-year-olds showing the highest coefficients, the 12-year-olds the next highest, and the 6-year-olds, the lowest. The rank order of the Full Scale correlations is also the same in both intelligence groups, but the order is not the same as for the Verbal Scale correlations. In this case, the 9-year-olds show the highest coefficient, but the 6-year-olds are second highest in this case, with the 12-year-olds last. With the exception of the fact that the 12-year-olds both show the lowest coefficients, there is no rank order agreement between the two intellectual groups for the Performance Scale correlations.

#### Defective Subjects

The failure to obtain a sufficient number of the more intelligent defectives (described in Chapter II) resulted in only five of this group achieving WISC Full Scale intelligence quotients. As a result no correlations were run in this sample. The actual results

will be discussed later, under a separate heading.

(b) Comparison of Mean IQs :

Superior Subjects

The comparison of the Mean intelligence quotients between the two tests for the Superior subjects are shown in Tables V, VI, and VII, (pp.29, 30 and 31).

In the Superior group, all the Mean WISC IQs are smaller than the Mean Binet IQs with which they are being compared.

In the Superior 6-year-old group, the Stanford-Binet IQs range from 110 to 154, a difference of 44 IQ points, the largest range in this age-intellectual group; the Binet standard deviation is slightly larger than those of the Wechsler Intelligence Scale for Children in this age-intellectual group. The WISC Verbal Mean IQ for the Superior 6-year-olds differs by only a small amount from the Binet Mean IQ for the same group and the difference is not significant; the range of IQs on the Verbal Scale in this age-intellectual category is from 105 to 131, a range of 26 IQ points; this scale also shows the smallest standard deviation in this age-intellectual group. The Mean Performance IQ for the Superior 6-year-olds differs from the Mean Binet IQ by the same amount as does the Verbal Scale; again, the difference is not significant. The IQs range from 106 to 138, a range of 32 IQ points, the largest range of any WISC scale in this age-intellectual group;

TABLE V

MEAN WISC VERBAL SCALE IQs COMPARED  
WITH MEAN STANFORD-BINET IQs

## Superior Subjects

Age	Stanford-Binet			WISC Verbal Scale				
	M. IQ	S.D.	S.E. <sub>M</sub>	M. IQ	S.D.	S.E. <sub>M</sub>	M. Diff.	t <sup>+</sup>
6 years	122.80	11.73	3.91	119.30	7.87	2.62	3.50	.90
9 years	137.90	17.00	5.66	125.20	11.86	3.95	12.70	5.16
12 years	131.90	8.75	2.92	113.20	8.07	2.69	18.70	6.65

<sup>+</sup>.05 = 2.26  
.01 = 3.25

TABLE VI

MEAN WISC PERFORMANCE SCALE IQs COMPARED  
WITH MEAN STANFORD-BINET IQs

Superior Subjects

Age	Stanford-Binet			WISC Performance Scale				
	M. IQ	S.D.	S.E. <sub>M</sub>	M. IQ	S.D.	S.E. <sub>M</sub>	M. Diff.	t <sup>+</sup>
6 years	122.80	11.73	3.91	119.30	10.56	3.52	3.50	.97
9 years	137.90	17.00	5.66	119.10	10.62	3.54	18.80	4.47
12 years	131.90	8.75	2.92	123.90	10.67	3.56	8.00	1.75

<sup>+</sup>.05 = 2.26  
.01 = 3.25

TABLE VII

MEAN WISC FULL SCALE IQs COMPARED  
WITH MEAN STANFORD-BINET IQs

Superior Subjects

Age	Stanford-Binet			WISC Full Scale				
	M. IQ	S.D.	S.E. <sub>M</sub>	M. IQ	S.D.	S.E. <sub>M</sub>	M. Diff.	t <sup>+</sup>
6 years	122.80	11.73	3.91	121.10	9.03	3.01	1.70	.49
9 years	137.90	17.00	5.66	124.40	11.00	3.66	13.50	4.86
12 years	131.90	8.75	2.92	120.20	7.33	2.44	11.70	3.74

<sup>+</sup>.05 = 2.26  
.01 = 3.25

the standard deviation is slightly larger than in the Verbal Scale. The Mean Full Scale IQ for the Superior 6-year-olds differs from the Mean Binet IQ by a much smaller amount than do the Verbal and Performance Scales; the difference is not significant. The IQs range from 108 to 138 points, a difference of 30 IQ points; the standard deviation is smaller than that of the Performance Scale, and larger than that of the Verbal Scale. All the WISC scale standard deviations in this age-intellectual group are smaller than the Binet standard deviation, but there is not a great difference between any of them.

In the group of Superior 9-year-olds, the Stanford-Binet IQs range from 119 to 168, a range of 49 IQ points, the largest in the entire study; this group of IQs also shows the largest standard deviation in the study; it shows a considerable difference from the next largest standard deviation. The Mean Verbal IQ in the Superior 9-year-old group shows the fourth greatest difference from the Mean Binet IQ in the entire study, although it is the smallest difference in this age-intellectual group; the difference is significant at the 1% level of confidence. The IQs range from 110 to 148, a difference of 38 points, and the largest WISC range in this age-intellectual group; it also shows the largest WISC standard deviation at this age-intellectual level. The Mean Performance IQ differs from the Mean Binet IQ by the largest amount in the study; the difference is significant at the 1% level of confidence. The IQs range from 104 to 136, a range of 32 points, the smallest range in this age-intellectual group; the



standard deviation is also the smallest in the age-intellectual group. The Mean Full Scale IQ for the Superior 9-year-old group shows the third largest difference from the Mean Binet IQ in the entire study; the difference is significant at the 1% level of confidence. The IQs range from 112 to 146, a difference of 34 points. The range of the sizes of the standard deviations is smaller in this age-intellectual group than for the Superior 6-year-old group, but the values are all larger.

In the Superior 12-year-old group, the Stanford-Binet IQs range from 118 to 146, a difference of 28 IQ points, the smallest Binet range in this intellectual group. The Superior 12-year-old group also shows the smallest Binet standard deviation in this intellectual group. The Mean Verbal Scale IQ shows the second largest difference from the Mean Binet IQ in the whole study, and the largest in this age-intellectual group; the difference is significant at the 1% level of confidence, showing the largest critical ratio obtained in the entire study. The IQs range from 104 to 130, a difference of 26 IQ points. The Mean Performance IQ shows the largest non-significant difference from the Mean Binet IQ in the entire study; it is the smallest Mean Difference in this age-intellectual group. The range of IQs is from 100 to 140, a difference of 40 IQ points, and the largest range in this age-intellectual group; the standard deviation is the largest in this same group. The Mean Full Scale IQ for the Superior 12-year old subjects differs from the Mean Binet IQ by the smallest

significant difference at this intellectual level; the difference is significant at the 1% level of confidence. The IQs range from 108 to 133, a range of 25 IQ points which is the smallest range for this intellectual group; the standard deviation is also the smallest for the intellectual group.

The 6-year level has the smallest Mean Differences within each WISC scale; in the Verbal Scale, the 12-year level shows the largest difference, with the 9-year level coming next. This order is reversed for the Performance and Full Scales. In the Stanford-Binet, the largest standard deviation occurs at the 9-year level, the smallest at the 12-year level, and the middle value at 6 years. This order holds also for the WISC Full Scale. In the WISC Verbal Scale the largest deviation occurs at the 9-year level, the next largest at 12 years, and the smallest at 6 years; in the Performance Scale, the largest deviation is at 12 years, the next largest at 9 years, and the smallest at 6 years. The range of size for these deviations is only .11 IQ points. In the Stanford-Binet, the largest Mean IQ occurs at 9 years, the next largest at 12 years, and the smallest at 6 years. In the WISC Full and Verbal Scales, the highest Mean IQ occurs at 9 years, the next largest at 6 years, and the smallest at 12 years. In the Performance Scale, the highest Mean IQ occurs at the 12-year level, the next largest at the 6-year level, and the smallest at 9 years.

#### Average Subjects

The comparison of the Mean intelligence quotients

between the two tests for the Average subjects are shown in Tables VIII, IX and X (pp.36, 37 and 38).

In the average 6-year-old group, the Stanford-Binet IQs range from 89 to 112, a difference of 23 IQ points, which is the smallest range in the entire study; the standard deviation is smaller than any in the Superior group, and is the smallest in this age-intellectual group. The Mean Verbal Scale IQ for the Average 6-year-olds does not differ significantly from the Mean Binet IQ, but it is still the largest difference at this age-intellectual level. The IQs range from 90 to 109, a difference of 19 IQ points, the smallest range in the study, and the standard deviation is the largest in this age-intellectual group. The Mean Performance Scale IQ differs from the Mean Binet IQ by the smallest amount of any of the WISC scales in this age-intellectual group; the difference is not significant. The IQs range from 89 to 114, making a difference of 25 IQ points; the standard deviation is the middle value deviation for the WISC scales at this age-intellectual level. The Mean Full Scale IQ differs from the Mean Binet IQ by a small amount which does not constitute a significant difference. The IQs range from 91 to 113, a difference of 22 IQ points, and the standard deviation is the smallest WISC deviation in this age-intellectual group. All the WISC Mean intelligence quotients at this level are smaller than the Mean Binet intelligence quotient.

At the average 9-year-old level, the Stanford-Binet IQs range from 84 to 113, a difference of 29 IQ points, and the standard

TABLE VIII

MEAN WISC VERBAL SCALE IQs COMPARED  
WITH MEAN STANFORD-BINET IQs

Average Subjects

Age	Stanford-Binet			WISC Verbal Scale				
	M. IQ	S.D.	S.E. <sub>M</sub>	M. IQ	S.D.	S.E. <sub>M</sub>	M. Diff.	t <sup>+</sup>
6 years	105.00	6.56	2.19	101.30	7.71	2.57	3.70	1.32
9 years	99.40	8.08	2.69	101.60	6.39	2.13	2.20	1.28
12 years	101.30	8.03	2.68	95.20	6.73	2.24	6.10	2.45

<sup>+</sup>.05 = 2.26  
.01 = 3.25

TABLE IX

MEAN WISC PERFORMANCE SCALE IQs COMPARED  
WITH MEAN STANFORD-BINET IQs

Average Subjects

Age	Stanford-Binet			WISC Performance Scale				
	M. IQ	S.D.	S.E. <sub>M</sub>	M. IQ	S.D.	S.E. <sub>M</sub>	M. Diff.	t <sup>+</sup>
6 years	105.00	6.56	2.19	102.00	7.24	2.41	3.00	1.33
9 years	99.40	8.08	2.69	98.20	7.11	2.37	1.20	.35
12 years	101.30	8.03	2.68	99.90	7.93	2.64	1.40	.43

<sup>+</sup>.05 = 2.26  
.01 = 3.25

TABLE X

MEAN WISC FULL SCALE IQs COMPARED  
WITH MEAN STANFORD-BINET IQs

Average Subjects

Age	Stanford-Binet			WISC Full Scale				
	M. IQ	S.D.	S.E. <sub>M</sub>	M. IQ	S.D.	S.E. <sub>M</sub>	M. Diff.	t <sup>+</sup>
6 years	105.00	6.56	2.19	101.60	6.59	2.20	3.40	1.55
9 years	99.40	8.08	2.69	99.90	5.11	1.70	.50	.27
12 years	101.30	8.03	2.68	97.10	6.69	2.23	4.20	1.63

<sup>+</sup>.05 = 2.26  
.01 = 3.25

deviation is the largest in this intellectual group. The Mean Difference between the Mean Verbal IQ for Average 9-year-olds and the Mean Binet IQ is not significant, but it is the largest in this age-intellectual group. The IQs range from 89 to 113, making a difference of 24 IQ points, the standard deviation is small. The Mean Performance IQ differs from the Mean Binet IQ by the second largest amount for this age-intellectual group; again, the difference is not significant. The IQs range from 89 to 110, the smallest range in this age-intellectual group; the standard deviation is the largest in this age-intellectual group. The Mean Full Scale IQ differs from the Mean Binet IQ by the smallest amount in the study; the difference is not significant. The IQs range from 88 to 109, making a difference of 21 IQ points, the same size as the range of the Performance Scale, at this age-intellectual level; the standard deviation is the smallest in the entire study. All the Mean Differences at this age-intellectual level are smaller than those at the Average 6-year-old level, and all the WISC scales show higher Mean intelligence quotients than does the Stanford-Binet.

In the Average 12-year-old group, the IQs range from 82 to 118, a difference of 36 IQ points, the largest range in this intellectual group; the standard deviation is only the second largest Binet deviation in this intellectual group, but it is the largest in the age-intellectual group. The Mean Verbal IQ differs from the Mean Binet IQ by the largest amount in this intellectual group, although the difference is not as large as the largest insignificant difference;

the difference is significant at the 5% level of confidence, and is the only significant difference in this intellectual group. The IQs range from 85 to 108, a difference of 23 IQ points. The Mean Performance IQ differs by the smallest amount in this age-intellectual group from the Mean Binet IQ; the difference is not significant. The IQs range from 89 to 110, a difference of 21 IQ points, the smallest range in this age-intellectual group; the standard deviation is the largest for this age-intellectual group. The Mean Full Scale IQ differs from the Mean Binet IQ by an amount which is not significant. The IQs range from 86 to 109, a difference of 23 IQ points; the standard deviation is the smallest for this age intellectual group. All the Mean WISC scale IQs are smaller than the Mean Binet IQ at this age-intellectual level.

In the Verbal and Full Scales, the largest differences occur at 12 years, then 6 years, then 9 years. In the Performance Scale, the largest difference is at age 6 years, then 12 years, then 9 years; the smallest difference is always at the 9-year level. The largest Verbal standard deviation occurred at 6 years, the next largest at 12 years, and the smallest at 9 years; in the Performance Scale, the largest standard deviation occurred at 12 years, the next largest at 6 years, and the smallest at 9 years; in the Full Scale, the order was the same. In all the WISC scales, the 9-year level shows the smallest standard deviation. In the Stanford-Binet at this intellectual level, the 9-year level shows the largest standard deviation; the



12-year level, the next largest; and the 6-year level, the smallest.

In the Stanford-Binet, the 6-year level shows the highest Mean IQ; the 12-year level, the next highest; and the 9-year level, the lowest. For the Verbal Scale, the highest Mean IQ occurs at 9 years; the next highest, at 6 years; and the lowest at 12 years. For the Performance Scale, the highest Mean IQ occurs at the 6-year level, the next highest at the 12-year level; and the lowest at the 9-year level. For the Full Scale, the highest Mean IQ occurred at the 6-year level, the next highest at the 9-year level, and the lowest at the 12-year level.

(c) The performance of the Defective Group :

Of the twelve Retarded 12-year-olds, all twelve obtained intelligence quotients on the Stanford-Binet; the IQs ranged from 28 to 68, a difference of 40 IQ points. Of these twelve, only three obtained WISC Full Scale IQs; these ranged from 55 to 67, a difference of 12 IQ points. Of the remaining nine, one subject achieved an IQ on both the Verbal and Performance Scales, and two subjects obtained IQs on the Verbal Scale alone. Of the remaining six subjects, one achieved no scaled score at all, and the others achieved scaled scores only on the Verbal and Full Scales.

In the group of Retarded 9-year olds, only five of the nine subjects achieved Stanford-Binet IQs; these IQs ranged from 26

to 60, a difference of 34 IQ points. Of these five, the two subjects, who achieved the highest Binet IQs for this group, also achieved Full Scale WISC IQs of 49 and 53. All the subjects in this age group achieved scaled scores on the WISC Full, Verbal and Performance Scales, because, due to the system of standard scores, it is possible for a subject at this, and lower age levels, to achieve scaled scores without having given a single correct response. However, two of the subjects did achieve scaled scores (but no IQs) because they had given some correct responses.

Of the group of four 6-year-old defectives, only one achieved a Stanford-Binet IQ (46) and none achieved a WISC Full Scale IQ. The three subjects who did not achieve Binet IQs, all achieved Verbal and Performance IQs, and the subject who achieved the Stanford-Binet IQ achieved a Verbal IQ but no Performance IQ, although he did obtain a scaled score on that section of the test. All the subjects achieved Full Scale scaled scores. The achievement of Verbal and Performance IQs by three of the subjects is due to the construction of the test, as outlined before, and no correct responses were actually given.

In the case of the 12-year-olds, the WISC Full Scale IQs were of a size comparable to those on the Binet, and this is true of all the Verbal and Performance IQs obtained at this age level. In the 9-year old group, the WISC Full Scale IQs are both smaller than the accompanying Binet IQs.

The actual scores obtained by these subjects are shown in Tables, XI, XII and XIII, (pp. 44, 45 and 46).

TABLE XI

PERFORMANCE OF 12-YEAR OLD MENTAL DEFECTIVES  
ON WISC VERBAL, PERFORMANCE, AND FULL SCALES,  
AND ON STANFORD-BINET

Subject Number	Stanford- Binet IQ	WISC					
		<u>Verbal</u>		<u>Performance</u>		<u>Full Scale</u>	
		Scaled Score	IQ	Scaled Score	IQ	Scaled Score	IQ
1	68	25	69	29	71	54	67
2	36	0	-	0	-	0	-
3	47	8	47	8	-	16	-
4	57	5	-	0	-	5	-
5	37	1	-	0	-	1	-
6	32	2	-	0	-	2	-
7	41	1	-	0	-	1	-
8	28	3	-	0	-	3	-
9	46	12	52	8	-	20	-
10	52	11	51	27	68	38	55
11	43	9	48	12	47	21	-
12	53	11	51	28	69	39	56

TABLE XII

PERFORMANCE OF 9-YEAR-OLD MENTAL DEFECTIVES  
ON WISC VERBAL, PERFORMANCE, AND FULL SCALES,  
AND ON STANFORD-BINET

Subject Number	Stanford- Binet IQ	WISC					
		<u>Verbal</u>		<u>Performance</u>		<u>Full Scale</u>	
		Scaled Score	IQ	Scaled Score	IQ	Scaled Score	IQ
1	26	2	-	1	-	3	-
2	+	2	-	5	-	7	-
3	+	2	-	2	-	4	-
4	60	14	55	21	60	35	53
5	52	12	52	18	55	30	49
6	+	2	-	1	-	3	-
7	41	3	-	7	-	10	-
8	25	2	-	1	-	3	-
9	+	2	-	1	-	3	-

<sup>+</sup>Impossible to compute an  
IQ because no basal age  
could be found.

TABLE XIII

PERFORMANCE OF 6-YEAR-OLD MENTAL DEFECTIVES  
ON WISC VERBAL, PERFORMANCE, AND FULL SCALES,  
AND ON STANFORD-BINET

Subject Number	Stanford- Binet IQ	WISC					
		<u>Verbal</u>		<u>Performance</u>		<u>Full Scale</u>	
		Scaled Score	IQ	Scaled Score	IQ	Scaled Score	IQ
1	+	11	51	12	47	23	-
2	+	11	51	12	47	23	-
3	+	11	51	12	47	23	-
4	46	12	52	9	-	21	-

+ Impossible to compute an IQ  
because no basal age could  
be found.

## CHAPTER IV

### DISCUSSION OF RESULTS

#### Theoretical Expectations and Obtained Results

As has been previously pointed out, since the Stanford-Binet has a standard deviation smaller at age 6 years, larger at 12 years, and approximately the same size at age 9 years, than that of the Wechsler Intelligence Scale for Children, there should be certain differences in the Mean IQs yielded by the two tests when testing intelligence at the extremes of the normal curve. At the 6-year level, for instance, the Mean Stanford-Binet IQ for superior subjects should be lower than the Mean WISC IQs; at the 12-year level, the Binet IQ for superior subjects should be higher than the Mean WISC intelligence quotients; and, at age 9 years, the Mean IQs on the two tests should be approximately the same.

These expectations are actually fulfilled at the Superior 12-year level where the Binet Mean IQ is considerably higher than the WISC Mean IQs, although the Mean Difference between the Binet and Performance Scale is not significant. The differences are far larger, however, than might be expected in terms of the sizes of the

standard deviations for the two scales. Furthermore, inspection of the differences obtained in the Superior groups at the other two levels tend to nullify the results of the Superior 12-year-old group of any significance as far as the theory about the standard deviations is concerned. Although none of the differences between the Mean Binet and the WISC Scale IQs at the 6-year level is significant, all the WISC scale Mean IQs are smaller than the Mean Binet IQ with which they are being compared. Moreover, in the Superior 9-year-old group, where the closest agreement should be expected, there occur three of the four largest Mean Differences in the study, all significant at the 1% level of confidence; again, the Mean Binet IQ is higher than any of the Mean WISC intelligence quotients.

In general, these results are fully substantiated by the size of the obtained standard deviations. With one exception, the Stanford-Binet standard deviations are larger than the deviations obtained by the same group on the WISC scales. The exception occurs in the Superior 12-year-old group where the Performance Scale standard deviation is larger than that of the Stanford Binet. It will be remembered that although the Mean Difference here was large, it was not significant. It is difficult to say, therefore, whether or not Wechsler's objections to the unusual sizes of the Stanford-Binet standard deviations at ages 6 and 12 years are justified on practical grounds. Within the limits of this study and compared with his own point scale with its set deviations, they certainly are not justified.



In general, the Average group appears to have achieved similar IQs on both the Stanford-Binet, and Wechsler Intelligence Scale for Children, Verbal, Performance and Full Scales. There is one exception to this trend, and that is when the 12-year-old Mean Binet IQ is compared with the Mean Verbal IQ. This difference is significant at the 5% level of confidence.

The nature of the obtained standard deviations is not always the same as those which occurred in the standardization groups of the two scales. It is rather interesting to note that in neither the Superior nor the Average groups do the Stanford-Binet standard deviations, obtained at the three age levels, rank in size in the same order, as do the standard deviations obtained with the original Binet standardization group. According to the latter, the 12-year-old group should show the largest standard deviation, the 9-year group the next largest, and the 6-year group the smallest. In the Superior group in this study, the 9-year group shows the largest standard deviation; the 6-year group the next largest; and the 12-year group the smallest. In the Average group, the 9-year group again shows the largest standard deviation, the 12-year group the next largest, and the 6-year group the smallest. On the other hand, within the results for each WISC scale of the Average group, and in the Superior group's Performance Scale, the standard deviations at all ages show considerable similarity, as would be expected by the fixed nature of the WISC standard deviations. Such similarity does not occur, however, in the results of the Superior

group for the Verbal and Performance Scales. The dissimilarity of these scales in this matter must be regarded as mere chance variations, since no explanation can be found in terms of unusual ranges. In only one instance, which occurs in the Superior 9-year-old Stanford-Binet results, are the obtained standard deviations as large as those reported by the authors of the two scales.

The fact that the size of the obtained Stanford-Binet deviations at the three age levels do not rank in the same order as do those in the original standardization group, is probably due to sampling errors, and could be explained by the differences in the range of obtained IQs. A sample of ten is much more subject to influence by a single extraordinary score than is a larger group. The larger size of the Binet standard deviations as compared with those of the WISC could be explained in terms of the point scale nature of the WISC with its set deviations, a matter to be discussed later in this section.

The smallness of the obtained standard deviations compared with those on the standardization groups of the two tests can probably be explained by the presence of factors intrinsic in the present experimental design. The limited size of the sample would make for a considerable restriction in the range of talent in any one group, that is, the homogeneous nature of the samples probably accounts best for the small size of the standard deviations; the ranges of the Average and Superior groups are necessarily smaller than those of the standardization groups which encompass not one, but all, intellectual

levels. The exception which occurs in the results of the Superior 9-year-old group on the Stanford-Binet, is probably the result of chance influence. The possible range in the Superior group is necessarily less restricted than in the Average group, and it so happens that this group contains two exceptionally high intelligence quotients of 164 and 168.

These differences between the two scales with regard to obtained Means and standard deviations are probably best explained in terms of differences in test construction. The study by Seashore, Wesman and Doppelt (14), described in the introduction, it will be recalled points out that the range of IQs on a point scale is quite arbitrary. Since the range of IQs and standard deviations on the WISC is quite narrow, extreme scores are not to be expected. Thus, the range of IQs, which it is possible for a subject to achieve, is much more restricted. The obtained results, i.e., lower Mean WISC IQs for the Superior subjects, and smaller WISC standard deviations is, therefore, to be expected on the basis of the construction of the test. The approximate equality of the Mean IQs in the Average group is also to be expected, since a difference in test construction, such as we are concerned with here, manifests itself only at the extremes of the normal curve.

Since the WISC does not yield an IQ below 45, only five of the defective populations achieved Full Scale IQs. The apparent equality between the IQs at the 12-year level, with a slight lowering of

the WISC IQs at age 9, is not in accordance with the varifying Binet standard deviations, or the restricted IQ range of the WISC. The sample is far too small, however, to permit any safe generalizations.

For reasons outlined in Chapter II and because it yields an IQ below 45, the Stanford-Binet does appear to be the preferable instrument at this level.

In terms of the many objections to the Mental Age concept, there is undoubtedly much to be said for the point scale, and a "conservative" one, such as the WISC which sets its scale so that extreme scores are rare, is useful because, for example, a Superior score is fairly certain to indicate superior ability. It is possible, however, that the "compression" of the values eliminates certain shadings of quantity, especially at that point between the average and the extremes. These shadings are obtainable on a scale like the Stanford-Binet, where the subject and his performance set the index to a far greater degree. Here the subject's age is set to the month, rather than within four months, and there is actually no IQ which he cannot attain. Whether the reliability and validity of the Binet, or of any intelligence test, warrants such precision is another question. One rather confusing characteristic manifested by the point scale nature of the WISC is the fact that, at the upper age and intelligence levels, a subject may make a perfect performance on a subtest and not achieve a perfect scaled score for it, or, even more, confusing, at the lower age and intellectual level, the subject may give no correct responses whatever, and still achieve a

scaled score.

Of the eighteen correlations computed, only five were significant, and all of these occurred at the 9-year level. It would appear then that, except at one age level, these two tests are not measuring the same thing. There are two major explanations as to why this may appear to be so.

The first one could be in terms of the motivational and intellectual development of the child during these years. A 6-year old child in Grade I does not have all the desirable attitudes and attentiveness more common in an older child; one encounters, in lesser degree, all the motivational problems of testing a very young child. As Anderson (1) points out, the younger the child, the less reliable the result; variation in motivation could account for the low correlations at this age. With a few exceptions, this problem is not a serious one in testing the 9-year-old; by this age, the child has been in school for three or four years and has, presumably, acquired the desirable attitudes and motivation. According to Thurstone (p.206, 18), a child of this age has no manifest special abilities due to a maturational factor, thus giving the impression of global intelligence; on the basis of this theory, the high correlations which occur at this age level would be expected. As the child grows older, he tends to develop specialized abilities; thus, the 12-year-old correlations are low due, presumably, to the fact that a good Performance Scale IQ does not always mean a good Verbal Scale IQ at this age level.

There are several objections to this explanation.

Although the correlations at the 9-year level are higher than at other ages, the Performance Scales in both intellectual groups correlate lowest at this age - evidence that a high IQ on one WISC scale does not necessarily mean a high IQ on the other. A study by Swineford (16) gives evidence contrary to the results suggested by this present investigation that there are special abilities at 9 or 12 years of age. Students in Grades VIII and IX (who would be older than 12 years of age) showed no change in factorial composition from one year to the next, and Swineford concludes that with increasing maturity, the general factor (presumably, whether factual or artifactual) plays a less important part.

The second explanation could be purely in terms of chance. The influence of chance factors in a sample of ten is bound to be much more distortive than in a larger sample. Presumably, the action of Probable Errors of the IQ within a small and homogeneous group tends to obscure any true relationship which might exist. In a large sample, a few reversals in rank order would be hidden by the general trend, whatever it might be .

Whatever the true explanation, keeping in mind the limited sample upon which this study was based, the two scales do not seem to be interchangeable. It does seem possible that a child given both of these tests might well yield widely different intelligence quotients on the respective tests.

The high correlations with large Mean differences within the Superior 9-year-old range are readily explainable. The high correlations indicate that the tests are certainly measuring the same ability in this group, and the differences in scores are attributable to differences in test construction. The WISC is a scale with a set standard deviation, in which the range has been comparatively restricted so that lower scores in this indicate much higher ability than the same score in the Stanford-Binet; for instance, an IQ of 130 amongst a group of Superior children, such as those encompassed by this study, is quite usual on the Stanford-Binet, but is much rarer on the WISC.

The dissimilarity of intelligence quotients yielded by the two tests is especially noticeable in the Superior 12-year-old group: correlations are all low and the Verbal and Full Scale IQs show a significant difference from the Mean Binet IQ at this age-intellectual level. Inspection of the actual scores reveals extreme differences and, with rare exceptions, these differences are in the direction of a lower WISC score. Such a situation is bound to be confusing in some degree to the clinician; he cannot be sure, for instance, when a subject receives a slightly-above-average IQ on the WISC, whether he would have made a comparable showing on the Stanford-Binet, or whether, as happened several times in this group of Superior 12-year-olds, he might have made a Superior showing in the Stanford-Binet. This latter does not constitute a fixed criterion by any means, but it certainly is a criterion of some sort, if only through long use, and one which cannot

be entirely ignored.

It is interesting to note that in both intellectual groups, the rank order of the three correlations (Binet with Verbal, Performance, and Full Scales) is the same for the same age, i.e., the rank order of the correlations is the same for the Superior 6-year-olds as for the Average 6-year-olds; the same for the Superior 9-year-olds as for the Average 9-year-olds; and the same for the Superior 12-year-olds as for the Average 12-year-olds. Any similarity in rank order of these correlations, within each intellectual group, is robbed of its significance by the fact that both intellectual groups display exactly the same rank order at each age. Furthermore, the highest, second highest, median, third lowest and lowest correlations, in both intellectual groups, occur in precisely the same correlations, i.e., with the same WISC scales (Verbal, Performance or Full) and at the same age levels. These results give some suggestion that age level influences much more the degree to which the Stanford-Binet and Wechsler Intelligence Scales for Children measure the same thing, than does intellectual level.

#### Obtained Results compared with the Literature

As has been previously pointed out, these results agree very largely with the predictions made by Seashore, Wesman and Doppelt (14). The findings also agree, in the main, with the results published by Frandsen and Higginson: the correlations at the 9-year level are high (sometimes slightly higher than those reported by Frandsen and



Higginson) with the Performance Scale correlating the lowest. This study did not find, however, that such high correlations occur also at other ages as is implied by the Frandsen and Higginson study, (5).

The results agree with the findings of Sartain (13), on the Stanford-Binet and Wechsler-Bellevue, that Superior subjects score higher on the Binet, although most of the differences found on this study are greater than the 5 IQ points suggested by him; similar findings were made by Goldfarb (6).

The results for the Retarded group are simply not adequate to permit a comparison of the findings with those of the literature. In agreement with Wechsler (p.157, 19), however, there does appear to be some trend toward the Performance intelligence quotient being higher than either the Verbal or Full Scale IQs.

The results of this study agree very closely with the findings of Halpern (9). She found that the highest correlations between the Stanford-Binet and Wechsler-Bellevue occurred at the extremes of the normal curve, and that, in spite of this fact, the greatest differences occurred with the Superior subjects and the least with the Defectives.

In general, this study agrees with the literature of the Stanford-Binet and Wechsler-Bellevue, in showing lower correlations between the Binet and WISC Performance Scale, than between the Binet and WISC Verbal or Full Scales.

### Consistent Trends emerging from the Data

The Superior 9-year-old will probably score much higher on the Stanford-Binet than on the WISC, but his position relative to the Mean is probably similar on both scales. The results for the Superior 12-year-old are a great deal more confusing; the two tests do not appear to be measuring the same phenomenon at this age-intellectual level. Much more research should be done at this, and surrounding, age-intellectual levels to ascertain whether or not the results of this investigation are due merely to chance influences, or whether they are due to some difference in subject matter in the two tests, or to the influence of some maturational factor.

One should also remember that the WISC does not yield an IQ below 45, and that, therefore, the Binet is much more useful in testing at this level of intelligence.

The fact that age influences the degree of correlation between the Binet and each of the three WISC scales, much more than does intellectual level, may be in agreement with some theories of general or special abilities and the influence of maturation. As was pointed out previously, however, the trend is not clearly towards either specific or general abilities.

### Suggestions for future Research

This study should be repeated to ascertain whether or not

the trends discerned in it are actual or merely the result of various chance influences; such a study would probably be rendered more reliable by the use of a larger sample. Since the results of this investigation suggest that the use of the WISC at very low levels of intelligence is not entirely satisfactory, future studies might utilize a less extreme group of defectives.

As outlined in a previous section, a future investigation might be directed toward ascertaining whether or not the wide differences in intelligence quotients, yielded by the two tests at the Superior 12-year-old level, are a manifestation of the development of special abilities. This purpose could, perhaps, be accomplished by testing above and below the 12-year age level, and noting consistent trends toward smaller correlations and wider significant differences with increasing age.

## SUMMARY

The purpose of this study was to investigate the relationship between IQs yielded by the 1937 Revised Stanford-Binet, Form L, and the Wechsler Intelligence Scale for Children (WISC) at different age levels and different levels of intelligence. The investigation endeavored to find out to what extent the Stanford-Binet correlates with each of the WISC Verbal, Performance and Full Scales at ages 6, 9 and 12 years, and with superior, average and defective children. It attempted to ascertain whether the Mean scores in this group show real differences in average IQs yielded by the two tests, and, if so, in which direction these differences occur.

The WISC is a point scale with a set standard deviation, while the size of the Stanford-Binet varies, being unusually small at age 6 years, and unusually large at age 12 years. It was hypothesized that at the extremes of the intelligence distribution at these ages, there should be a difference between the Mean IQs of the two scales in the direction corresponding to the size of the standard deviation at the age in question. To insure that extreme scores would occur, the sample included superior subjects and mental defectives. Besides the 6 and 12-year-olds necessary to test this hypothesis, the sample also

included 9-year-olds, whose Stanford-Binet standard deviation approximates that of the WISC. These 9-year-olds, plus a sampling of average children, were included for purposes of control and comparison but also for their own research value.

The sample was composed of eighty-five subjects - ten children in each age category in the Superior and Average groups; and in the group of Defectives, fourteen 12-year-olds, nine 9-year-olds, and four 6-year olds. The Average and Superior subjects were obtained in the elementary schools throughout Greater Vancouver, and were chosen on the basis of past records of ability and achievement. The group of Defectives was obtained from the Woodlands School, New Westminster, B.C., and from the Rainier State School, Buckley, Washington, U.S.A.: all the feeble-minded subjects had been diagnosed as "familial" defectives.

The positive correlations which occurred may be summarized as follows :

1. In the 9-year-old Superior group, the Stanford-Binet IQ correlated significantly with -
  - (a) the WISC Verbal IQ at the 1% level of confidence;
  - (b) the WISC Performance Scale IQ at the 5% level of confidence;
  - (c) the WISC Full Scale IQ at the 1% level of confidence.
2. In the 9-year-old Average group, the Stanford-Binet IQ correlated significantly with -

- (a) the WISC Verbal Scale IQ at the 1% level of confidence;
- (b) the WISC Full Scale IQ at the 5% level of confidence.

Significant differences between the Mean IQs of the two tests may be summarized as follows :

1. In the group of Superior 9-year-olds, the Stanford-Binet IQs were significantly higher (at the 1% level of confidence) than the WISC Verbal, Performance and Full Scale IQs.
2. In the group of Superior 12-year-olds, the Stanford-Binet IQs were significantly higher at the 1% level of confidence for the WISC Full and Verbal Scale IQs.
3. In the group of Average 12-year-olds, the Stanford-Binet is significantly higher at the 5% level of confidence than the WISC Verbal Scale IQ.

The major conclusions of this study are :

1. The obtained results are in essential agreement with the studies comparing the Wechsler adult scale and the Stanford-Binet.
2. There seems to be a consistent tendency in this study and others reviewed previously, toward lower correlations between the Stanford-Binet and WISC Performance Scale, than between the Stanford-Binet and WISC Verbal and Full Scales.
3. There seems to be no support for the hypothesis that the differences between the Mean Binet and

Mean WISC IQs at the Superior level will differ in direction according to the size of the Stanford-Binet at the age level in question.

4. The WISC appears to be an unsatisfactory test for the markedly retarded children. Both in terms of construction and interest value to subjects, the Stanford-Binet seems to be a better scale for the measurement of the lower levels of intelligence.
5. Keeping in mind the limited sample upon which this study was based, the two scales do not seem to be interchangeable. The practical import of this conclusion is that clinicians, social workers, psychiatrists, school teachers, and so on, should be fully aware that the child given both tests may well yield widely different IQs on the respective tests.

## BIBLIOGRAPHY

1. ANDERSON, John E. The limitations of infant and preschool tests in the measurement of intelligence. J. Psychol., 1939, 8, 351-379.
2. BALINSKY, B., ISRAEL, H., and WECHSLER, D. Relative effectiveness of the Stanford-Binet and Bellevue intelligence scales in diagnosing mental deficiency. Amer. J. Orthopsychiat., 1939, 9, 798-801.
3. BENTON, Arthur L., WEIDER, Arthur, and BLEAUVELT, Jean. Performance of adult patients on the Bellevue intelligence scales and revised Stanford-Binet. Psychiat. Quart., 1941, 15, 802-806.
4. BURT, Cyril. Critical notice of Thurstone's "Multiple Factor Analysis." Brit. J. educ. Psychol., 1947, 17, 163-169.
5. FRANDSEN, Arden N., and HIGGINSON, Jay B. The Stanford-Binet and the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 236-238.
6. GOLDFARB, William. Adolescent performance in the Wechsler-Bellevue intelligence scale and the revised Stanford-Binet examination, Form L. J. educ. Psychol., 1944, 35, 503-507.
7. GUERTIN, Wilson H. Mental growth in pseudo-feble-mindedness. J. clin. Psychol., 1949, 5, 414-418.
8. GUILFORD, J. P. Human abilities. Psychol. Rev., 1940, 47, 367-394.



9. HALPERN, Florence. A comparison of the revised Stanford-Binet, Form L, and the Bellevue adult intelligence test as clinical instruments. Psychiat. Quart. Suppl., 1942, 16, 206-211.
10. KUTASH, S. B. A comparison of the Wechsler-Bellevue and the revised Stanford-Binet scales for adult defective delinquents. Psychiat. Quart., 1945, 19, 677-685.
11. MITCHELL, M. Performance of mental hospital patients on the Wechsler-Bellevue and the revised Stanford-Binet, Form L. J. educ. Psychol., 1942, 33, 538-545.
12. RABIN, Albert I., and GUERTIN, Wilson H. Research with the Wechsler-Bellevue test: 1945-50. Psychol. Bull., 1951, 48, 211-248.
13. SARTAIN A. A comparison of the new revised Stanford-Binet, the Bellevue scale and certain group tests of intelligence. J. soc. Psychol., 1946, 23, 237-239.
14. SEASHORE, Harold, WESMAN, Alexander, and DOPPELT, Jerome. The standardization of the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1950, 14, 99-110.
15. SPEARMAN, C. Theory of general factor. Brit. J. Psychol., 1946, 36, 117-131.
16. SWINEFORD, Frances. Growth in the general and verbal bi-factor from Grade VII to Grade IX. J. educ. Psychol., 1947, 38, 257-272.
17. TERMAN, Lewis M., and MERRILL, Maud A. Measuring Intelligence. London: George G. Harrap and Co. Ltd., 1937.
18. THURSTONE, L. L. Current issues in factor analysis. Psychol. Bull., 1940, 37, 189-237.
19. WECHSLER, David. The Measurement of Adult Intelligence. Baltimore: The Williams and Wilkins Company, 1944.

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