1957

A Study of the Reliability of the Wechsler Adult Intelligence Scale with Eighteen and Nineteen Year Olds

Fred Donnel Whelan
Loyola University Chicago

A STUDY OF THE RELIABILITY OF THE WECHSLER ADULT INTELLIGENCE SCALE WITH EIGHTEEN AND NINETEEN YEAR OLDS

by

Fred D. Whelan

A Dissertation Submitted to the Faculty of the Graduate School of Loyola University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

June

1957

LIFE

Fred Donnel Whelan was born in Villa Ridge, Illinois, August 10, 1921.

Upon graduating from high school he attended LaSalle-Peru-Oglesby Junior College, LaSalle, Illinois, until he enlisted in the Armed Forces in September, 1942. After continental and foreign service he was honorably discharged April 1, 1946.

In February, 1949, he was graduated from Loyola University, Chicago, Illinois, with the degree Bachelor of Philosophy.

From 1948 to 1950, he served as a psychological trainee in the Loyola Center for Psychological Services and Child Guidance. In 1950, he accepted the position of psychologist with the Shiel Guidance Service, Catholic Youth Organization, Chicago, Illinois.

In June, 1951, he received the degree of Master of Arts from Loyola University, Chicago, Illinois. Since November, 1954, he has been employed as a supervising psychologist in the Guidance Center, Catholic Charities of Chicago.

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION	1
II.	RELATED LITERATURE	5
	A. Criticism of the Wechsler-Bellevue Scale	5
	B. Reliability studies of the Wechsler-Bellevue Scale	7
	C. Research on the Wechsler Adult Intelligence Scale	13
	D. Comparisons of the Wechsler Adult Intelligence Scale and Wechsler-Bellevue Intelligence Scale with the Stanford-Binet Intelligence Scale	29
III.	DESIGN OF RESEARCH	36
	A. Selection and Procurement of the Sample	36
	1. Criteria for selection of subjects	36
	2. Procedures in obtaining Sample	38
	3. Response of persons contacted	40
	4. Difficulties encountered in obtaining subjects	41
	5. Scoring of the Wechsler Adult and Stanford-Binet Intelligence Scales	42
	B. Statistical Analysis of the Wechsler Adult Intelligence Scale and Stanford-Binet Scale	43
	1. Intra-reliability of the subtests and IQ scales of the Wechsler Adult Intelligence Scale	43
	2. Inter-reliability of the Wechsler Adult and Stanford Binet Intelligence Scales	48
IV.	ANALYSIS OF RESULTS AND INTERPRETATION	49
	A. Representativeness of the sample	49
	B. Reasons for using the Split-half method of estimating reliability	52
	C. Intra-reliability of the subtests and IQ scales of the Wechsler Adult Intelligence Scale	53
	D. Inter-reliability of the Wechsler Adult and Stanford-Binet Intelligence Scales	56
	E. Distribution of Full Scale IQ's of the Wechsler Adult Intelligence Scale and IQ's of the Stanford-Binet Intelligence Scale	59

CHAPTER Page

V. SUMMARY AND CONCLUSIONS 61

BIBLIOGRAPHY 64
APPENDIX. 68

ACKNOWLEDGEMENTS

Expressions of appreciation are due to many individuals whose contributions made this study possible. The writer is especially indebted to Frank J. Kobler, Ph.D., Professor of Clinical Psychology, Loyola University, under whose direction this study was completed, for his constant interest, encouragement, and helpful criticisms. Grateful acknowledgement must also be made to Ralph D. Bergen, M.D., Director of the Guidance Center of Catholic Charities, whose understanding and patience enabled the writer to devote many working hours to the testing of subjects and to writing this paper. The writer also wishes to express his appreciation to the many teachers and employers who provided suitable subjects, and to the one-hundred subjects who gave so generously of their time without thought of reward.

Fred D. Whelan

CHAPTER I

INTRODUCTION

In the early years of the twentieth century, mental tests began to be used in the classroom for the purpose of obtaining an objective evaluation of a child's intelligence. In the intervening years since that time, the concept of measuring intelligence has pervaded many phases of American life. Today, the school teacher as well as the personnel manager relies upon the findings of mental tests to aid them to perform more effectively in their fields.

One of the major reasons why intelligence tests have become so widely used has been because their results have been found valid and reliable. However, before a test can be regarded as an accurate measure it must be proven in the research situation. Whenever a new test appears it is subjected to experimentation on population samples other than those of the original standardization group. It is through such research that a test is either accepted or rejected as a reliable instrument of measurement.

One of the most recently published individual intelligence tests is the Wechsler Adult Intelligence Scale (51).[1] This scale is essentially an extension and modification of Form I of the original Wechsler-Bellevue

---

[1]The abbreviation, WAIS, will be used throughout to indicate the Wechsler Adult Intelligence Scale.

Intelligence Scale published in 1939 (50).[2] This latter test was specifically designed for the purpose of measuring adult intelligence. Such a test was a needed instrument in the field of psychological testing because the majority of intelligence tests in use at that time had been standardized on and constructed for the purpose of measuring the intelligence of children. These tests, such as the 1916 and 1937 revision of the Stanford-Binet Intelligence Scale, were not suited for use with adults because of test content, the inapplicability of mental age norms to adults, and the fact that standardization samples included few adults.

Since publication, the Wechsler-Bellevue Scale became one of the most widely used intelligence tests in the United States. A recent survey reported that only the Stanford-Binet Scale continued to have greater usage (31). The fact that this scale fulfilled an urgent and long-standing need appears to account for this ready acceptance.

There are, however, several telling criticisms that can be directed against this scale even though it has gained wide popularity. The significance of these criticisms led to the publication of the WAIS in 1955 by The Psychological Corporation. The basic structure of this new scale remains the same as that of its predecessor, except for the inclusion of the Vocabulary subtest as a regular rather than an alternate subtest. The content of the subtests has been revised, some test materials have been redesigned, and the directions and scoring have become more precise and generally simpler.

---

[2]The abbreviation, W-B, will be used throughout to indicate the Wechsler-Bellevue Intelligence Scale, Form I.

The probability that this new test will have widespread use causes the writer to believe that it will be practical to have greater information than is presently available regarding this scale.

The primary purpose of this study was to determine the intrareliability of the WAIS. It is necessary not only to know the reliability of a test on the standardization group, but equally or more important to know the reliability of a test when administered to many different population samples. The test that measures consistently from sample to sample is the more valued instrument because the psychologist can depend upon the results of such a test. Thus, since the WAIS probably will be used widely, it is important to have reliability measures from various samples of the population.

A secondary purpose of this study was to compare the results of the WAIS with those of the Revised Stanford-Binet, Form L.[3] This purpose is of merit because the clinical usefulness of a test is enhanced when there is some knowledge of its relationship to similar measures. It is essential to know the degree of relationship between two measurements in order to validly interpret the results of each. Such information becomes especially important when retesting is necessary. In most instances, the same test form cannot be used in retesting because during the second administration the testee would probably get appreciably higher scores on timed items and those problems involving simple learning. It is expected that the SBL will continue to be used extensively and that the WAIS will eventually have similar usage. Information regarding the

---

[3]The abbreviation, SBL, will be used throughout to indicate the Revised Stanford-Binet, Form L.

comparability of these two measures would be of practical value to the clinician.

It is hypothesized that if the WAIS is a reliable instrument of measurement, the reliability measures obtained in this study will be comparable with those of the standardization group and similar studies. Furthermore, it will be interesting and profitable to make a comparison between the IQs obtained from the WAIS and SBL.

CHAPTER II

RELATED LITERATURE

A. Criticism of the Wechsler-Bellevue Scale.

Most writers critically evaluating the W-B, write of the need for a more representative sampling, especially with regard to geographical distribution, equal sampling of both sexes, and white and non-white subjects (2, 18, 34, 43, 48). These general criticisms have been met in the standardization of the WAIS. Equal numbers of men and women, both white and non-white, were selected and tested in twenty-four testing centers throughout the four major geographical regions of the United States.

In addition to these criticisms there are numerous specific objections to the W-B. For example, Gurvitz (20) states that Wechsler was careless regarding changes in subtest restandardization. He noted that in the second edition of the manual, Wechsler permitted the subject to have fifteen seconds in which to respond to each item of the Picture Completion subtest. In the third edition the subject was permitted "15" to 20" per picture" (50, p. 181). There was also a similar change in instructions noted for the Object Assembly subtest, but no change in the scoring for either the Object Assembly or Picture Completion subtests.

Other psychologists apparently agree with Gurvitz for Kitzinger and Blumberg (28) published a supplementary guide for administering and scoring the W-B. It included reasonable descriptions of subtest rationale and presented

definite scoring principles and examples.

Several studies have appeared wherein criticism was directed against the item order of certain subtests of the W-B. This aspect was first investigated by Jastak (26). He reviewed 1176 records and reported that by revising Wechsler's original item order he was able to obtain higher scores on Information, Arithmetic, Similarities, Vocabulary, Picture Completion, Block Design, and Object Assembly subtests. Norman (37) reported similar findings with a group of 153 subjects with IQ's of 120 or better. Mech (35), using high school students and Russell (44), using neuropsychiatric patients, also report higher scores on particular subtests when using a revised sequence of items. These writers seem to imply that a suitable item order should be found that will apply to any group of subjects. Guertin, et al, (17) appropriately comment, however, in their review of research with the W-B, that different orders of difficulty found with small restricted portions of the whole population should not be used to challenge the original order. Wechsler, one can assume, attempted to establish an item order based on a representative population rather than any particular group. Most subtests of the WAIS reflect revision of item order and addition of new items. Future research will determine whether this revised order is more applicable to general as well as specific population groups.

Anastasi (2) mentioned the preceeding criticisms and added that the reliabilities of some W-B subtests were too low to permit them to be used for clinical diagnosis based on scatter analysis. This comment coincides with the majority of research on this aspect. Guertin, et al, in reviewing such research, comment that the findings are inconclusive and that, "patterns

suitable for clinical use will not be forthcoming until methodological improvements appear" (17, p. 251).

Anastasi was also critical of the fact that the age groups in the original W-B standardization sample were not equated as to education. The younger subjects, on the average, had received more education than older subjects. This objection has been met in the WAIS standardization as all age groups are equally represented according to educational level. This writer was also of the opinion that the level of difficulty of certain W-B items, especially in the Information and Vocabulary subtests, probably had been affected by world events. For example, the item "harakiri" in the Vocabulary subtest has become easier for many adults because the incidence of this act in Japan was so widely publicized during and after World War II. Revision of many of the subtests in the WAIS indicates that Wechsler recognized this weakness and has attempted to reduce it in the new scale.

B. Reliability Studies of the Wechsler-Bellevue Scale.

Considering the extent to which the W-B is used, the number of studies investigating reliability of the scale is limited (42, 17). Most investigators (10, 23, 34, 35) have used the retest method of computing reliability and the majority of these studies have utilized psychiatric populations. Rabin and Guertin (43) object, however, to the use of such samples for they believe that the results are inconclusive because retest reliabilities have been influenced by clinical changes and practice effects.

The following two studies illustrate these criticisms. Hamister (23) reported on a study where the W-B was administered twice to fifty-three neuropsychiatric patients. The diagnoses of these patients embraced the major

psychiatric entities although the larger number, thirty-four, were schizophrenic. The retest was administered after one week to thirty-three patients and after one month to the remaining twenty. Total group mean age was 29.26. There were no significant differences found between the means of the subtests on test and retest. There was, however, a highly significant mean difference of 10.75 weighted score points between Full Scale scores on test and retest. One can presume that clinical changes in the patients in addition to practice effects contributed to this significant difference.

A second study using the retest method was that of Derner, et al, (10). The intelligence range for the 158 normal subjects employed in this research was from dull normal to superior. Sixty subjects were retested after one week; sixty after four weeks, and thirty-eight after six months. The writers found practice effects to be significant on Full Scale IQ's for all three groups. They noted, however, that the effect diminished as the time interval lengthened between tests.

Other studies have used the split-half method of estimating reliability. Concern with clinical changes or practice effects is reduced if this method is utilized and in the following studies this technique was employed.

Webb and de Hann (49) state that they believe the split-half technique should be used not only because of clinical changes and practice effect, but also because the W-B is usually administered in one test period. In other words, a single administration is the usual manner in which a test is employed and reliability should be obtained in the same manner, not through an irregular manner such as retesting.

They administered the W-B to fifty female paranoid schizophrenics and to

fifty normal females similar in age and IQ. Split-half reliabilities, corrected by the Spearman-Brown formula, were obtained on all subtests except the Digit Symbol subtest.

Table 1

Split-Half Reliabilities of W-B Subtests

(Webb and De Haan)

Subtest	Normals	Schizophrenics
	N=50	N=50
Information	.82	.84
Comprehension	.53	.73
Digit Span	.44	.54
Arithmetic	.82	.81
Similarities	.59	.73
Vocabulary	.94	.91
Picture Arrangement	.29	.63
Picture Completion	.42	.81
Block Design	.76	.86
Object Assembly	.46	.64

The obtained reliabilities, shown in Table 1, indicate higher split-half coefficients for the paranoid schizophrenic group than for the normal group in all Performance Scale subtests and in three of the Verbal Scale subtests. Only two differences between the coefficients, however, were significant; one at the one per cent level (Picture Arrangement) and one at the five per cent level (Picture Completion). The authors concluded that there was a difference

between the performance of normals and schizophrenics on the Performance Scale of the W-B and that higher reliabilities were to be obtained on psychiatric patients than on normals.

In reporting their results, they state that although they correlated Digits Forward with Digits Backward to obtain a reliability coefficient, as Wechsler did, it did not mean that they presumed that Digits Forward and Digits Backward measure the same ability. Apparently they believe that there is a difference as to what is measured by the two parts of this subtest. Whether or not Wechsler assumes that there is a real difference, it must be presumed that he considers the two parts to measure the same ability or he would have estimated the reliability coefficient in some other manner.

Helmick (24), in a critical review of Webb and de Haan's results, disagreed with their conclusions. He pointed out that since the higher reliability coefficients were dependent upon the variability in the groups, the conclusions of Webb and de Haan were unwarranted. He recalculated the data of their study and found that, when the respective variabilities were allowed for, the magnitudes of the reliability coefficients were about the same for the schizophrenic and the normal groups.

Botwinick (5) obtained further knowledge of W-B split-half reliabilities by making a comparison between the subtest reliabilities computed by Webb and de Haan (49) for their "normal" group and the subtest reliabilities of a "normal" but older group, and also between those of a matched group comprised of hospitalized patients diagnosed as "senile psychosis" and "psychosis with cerebral arteriosclerosis."

The mean age of Webb and de Haan's "young normal" group was 37.6 years.

These subjects were fifty females with a mean IQ of 97.8. The mean age of Botwinick's "old normal" group was 64.3 years. Fifty males and females, with a mean IQ of 100.8, comprised this group. The "senile" group consisted of thirty-one subjects both male and female. Mean age or IQ was not reported for this latter group.

Table 2

AGE DIFFERENCES IN RELIABILITY AND STANDARD ERROR OF MEASUREMENT OF W-B SUBTESTS (Botwinick)

TEST	Young Normal N=50		Old Normal N=50		Senile N=31	
	r.	SE_m	r.	SE_m	r.	SE_m
Vocabulary	.94	1.37	.96	1.48	.94	1.69
Arithmetic	.82	1.06	.70	1.11	.92	0.85
Information	.82	1.58	.90	1.21	.93	1.59
Block Design	.76	2.90	.84	2.51	.89	2.14
Similarities	.74	1.80	.89	1.47	.83	1.43
Comprehension	.53	1.96	.35	2.22	.69	1.89
Object Assembly	.46	1.90	.62	2.50	.66	2.82
Digit Span	.44	1.45	.76	1.07	.51	1.44
Picture Completion	.42	1.48	.83	1.25	.82	1.15
Picture Arrangement	.29	2.23	.55	1.57	.87	1.18

The coefficient of reliabilities and standard errors of measurements obtained by the split-half method are presented in Table 2. Comparisons in reliability coefficients between "young normal" and "old normal" groups revealed a significant difference at the one per cent level on only one subtest:

Picture Completion. Comparisons in reliability coefficients between the "old normal" and "senile" groups indicated a significant difference at the one per cent level between Arithmetic and Picture Arrangement subtests.

Botwinick concluded that his results were comparable with those of other reliability studies regardless of whether the split-half or retest method was employed.

Gilhooly (13), reports split-half reliabilities for four verbal tests of the W-B: Information .79, Comprehension .56, Similarities .77, and Vocabulary .94. The sample consisted of 122 male, World War II veterans, nineteen to thirty-five years of age, who were hospitalized for psychoneurosis. These reliabilities are comparable with those reported by Webb and de Haan (49) on a similarly aged normal group.

The reliability coefficients reported in each of these studies are so low on some of the subtests that the stability of the measures is highly questionable and appear to be too indicative of error to warrant confidence in making diagnostic evaluations based on scatter analysis. This is especially true of most performance subtests and the Digit Span, Comprehension, and Similarities subtests of the Verbal Scale.

It was disconcerting to discover that so few reliability studies have been done on the W-B when one considers the extent to which this test is used. Also, that none of the studies presented had compared their results with those of the standardization study in order to test the stability of their findings. It would seem to this writer that it would be important to establish the stability of a test on normal groups before determining reliabilities of atypical groups. When the reliability studies are done on the latter groups, one can never be

certain of whether the obtained reliabilities reflect true stability or whether they reflect, in part, the effects of clinical symptoms. The writer finds it difficult to accept the reliabilities reported by these studies because suitable criteria, the reliability coefficients of normal groups, are not available for comparison.

C. Research on the WAIS.

The WAIS is essentially an extension and modification of the original W-B (51). A number of changes have been made in the subtests and in the directions for administering and scoring. The basic structure remains the same, however, and those who are familiar with the W-B will find that many items of that scale have been retained in the WAIS. The following is a summary of the item changes within each subtest.

Information: The number of items in this subtest has been increased from twenty-five to twenty-nine. Sixteen items appeared in the original scale and thirteen are new. As in the W-B, one point is received for each correct answer.

Comprehension: The number of items is now fourteen as opposed to twelve in the W-B. Eight of the W-B items have been retained. The testee receives zero, one, or two points depending upon the correctness and quality of his response. Two items have been added at the lower end of this subtest which permits more valid testing of limited or borderline intelligence.

Arithmetic: Five items have retained from the W-B and nine items added for a total of fourteen items. Here also, two items have been added at the lower end of the subtest to permit more valid testing of low intelligence. In the W-B, time credit was given on the last two problems which were read aloud by the testee. The last four items of the WAIS subtest have time credits and all

problems are read by the examiner.

Similarities: The revised Similarities subtest consists of thirteen items ten of which were retained from the original scale. Scoring remains the same as in the W-B.

Digit Span: This subtest is identical in content to that of the W-B.

Vocabulary: This subtest is now included as a regular rather than an alternate subtest. The forty items are all new. The words are more familiar to the average and above average individual, whereas in the W-B the words at the higher end of the scale were familiar to few individuals. It is expected that even though testees may give totally incorrect answers, they will less apt to be aware of failure and test rapport will be easier to maintain. As on the W-B, the testee obtains zero, one or two points depending upon the quality and correctness of his response.

Digit Symbol: The number of items have been increased from sixty-seven to ninety. The symbols remain the same, except that the reversed "N", which resulted in many half-scores on the W-B, has been replaced by an upside-down "T". Half-scores are no longer obtainable. One point is received for each symbol drawn correctly.

Picture Completion: Twenty-one items comprise this subtest, eleven of which have been retained from the original scale. Certain ambiguous items, such as the picture of the man without a tie, have been eliminated. Scoring is the same as on the W-B.

Block Design: This subtest consists of ten items, seven of which appeared in the W-B. The number of blocks has been reduced from sixteen to nine. The color of the blocks are now red, white, and red-and-white; the blue, yellow

and blue-and-yellow of the W-B blocks have been discarded. The testee now receives credit for the two demonstration items which was not true of the original subtest. Bonus credits for time are no longer permitted until the testee reaches the more difficult designs, whereas on the W-B they were given throughout the subtest. Basic score is four points, as opposed to three on the W-B, and bonus points of one or two can be obtained on the last four items.

Picture Arrangement: This subtest is comprised of eight items. Six are retained from the W-B. Credit is now received for the demonstration item which was not true of the original subtest. Basic score is four points, opposed to two and three points on the W-B, and one or two bonus points may be earned if the last two items are completed correctly within a specified time period.

Object Assembly: The three items appearing in the original subtest have been retained and one new item has been added. The size of the Profile Assembly item has been reduced considerably permitting greater ease in presentation. Basic score is different for each item and additional points may be earned for successful completion within particular time limits.

The writer presumes that the subtests of the WAIS were lengthened for the purpose of increasing their reliability. Whether or not this has been accomplished will be determined by the results of future research. From the viewpoint of the clinician whose time is usually limited, the lengthening of the subtests requires longer administration time of from ten to twenty minutes. If future research on the W-B and WAIS reflects comparable testee performance, IQ results, and subtest reliability, it will be questionable whether the longer administration time of the WAIS will warrant it being used instead of the W-B.

The comparatively recent appearance of the WAIS and the publication lag

in the printing of journal articles are probable reasons for the paucity of research studies that have thus far appeared on the scale.

The only available material on reliability of the WAIS subtests and IQ's is the original standardization data done by Wechsler (51, pp. 12-13). These data consist of the reliability coefficients and standard errors of measurement on each subtest and IQ scales. The reliability coefficient for each subtest, except Digit Span and Digit Symbol, was determined by computing the correlation between scores on odd and even items and correcting the coefficient for full length of the test by the Spearman-Brown formula. The reliability coefficient for the Digit Span subtest was determined by obtaining the correlation coefficient between Digits Forward and Digits Backward and correcting this coefficient for full length of the test.

A special study was developed by Wechsler to estimate the reliability of the Digit Symbol subtest. The Digit Symbol subtest of the WAIS and the Digit Symbol subtest of the Wechsler-Bellevue Intelligence Scale, Form II, were administered to 132 female applicants to a school of nursing. The ages ranged from sixteen to twenty-four. The mean score on the WAIS Digit Symbol was 53.8 and the standard deviation was 10.8. The correlation coefficient between the two Digit Symbol subtests was .88. Of the 132 females, a subgroup aged eighteen and nineteen obtained a mean score of 53.6 on the WAIS Digit Symbol test and a standard deviation of 10.8. The correlation coefficient was again .88. The group of eighteen - nineteen year old females in the standardization sample obtained a mean score on the WAIS Digit Symbol of 56.4 and a standard deviation of 13.2. The reliability coefficient for the eighteen - nineteen standardization group was then estimated by correcting the obtained coefficient

of .88 for the difference between standard deviations of the group of nurses and the standardization group. The eighteen - nineteen year group was the only age group of the standardization sample for which Wechsler reported a reliability coefficient on this subtest.

The reliability coefficients for the Verbal, Performance, and Full Scale IQ's were estimated from the formula for the correlation between two sums of equally weighted scores (28, p. 396).

The obtained reliabilities and standard errors of measurement for three age groups are presented in Table 3. These three groups were those selected by Wechsler as being representative of the age ranges included in the standardization sample.

According to these reliabilities, some subtests of the WAIS, in spite of changes, remain unstable. This appears especially true of the Picture Arrangement and Picture Completion subtests. It will be noted that the Vocabulary subtest, the "old reliable" of test constructors, continues to be the most stable measure. It is expected that Wechsler, in his forthcoming book concerning the WAIS, will continue to promote diagnosis based on scatter analysis and the prediction of individual abilities. The low reliabilities of some subtests will require the clinician to be extremely cautious in this respect. A score on a subtest of low reliability may be due to chance or it may reflect the true ability of the testee. The lower the reliability coefficient the greater the probability that the score will be due to chance. Thus, interpretations made from comparisons between differences in scores on subtests of low reliability may be quite erroneous. This is also true in predicting ability from an individual's score on one subtest.

Table 3

RELIABILITY COEFFICIENTS AND STANDARD ERRORS OF MEASUREMENT*

OF THE WAIS (Wechsler)

Test	Age 18-19 N=200		Age 25-34 N=300		Age 45-54 N=300	
	r	SE_m	r	SE_m	r	SE_m
Information	.91	.88	.91	.86	.92	.87
Comprehension	.79	1.36	.77	1.45	.79	1.47
Arithmetic	.79	1.38	.81	1.35	.86	1.23
Similarities	.87	1.11	.85	1.15	.85	1.32
Digit Span	.71	1.63	.66	1.75	.66	1.74
Vocabulary	.94	.69	.95	.67	.96	.67
Verbal IQ	.96	3.00	.96	3.00	.96	3.00
Digit Symbol	.92	.85				
Picture Completion	.82	1.18	.85	1.14	.83	1.15
Block Design	.86	1.16	.83	1.29	.82	1.15
Picture Arrangement	.66	1.71	.60	1.73	.74	1.39
Object Assembly	.65	1.65	.68	1.66	.71	1.59
Performance IQ	.93	3.97	.93	3.97	.94	3.67
Full Scale IQ	.97	2.60	.97	2.60	.96	2.60

*The SE_m is in Scaled Score units for the tests and in IQ units for the Verbal, Performance, and Full Scale IQs.

The WAIS, as was stated above, is not a new scale but actually a modification of the W-B that was developed in order to meet certain limitations of its predecessor. The consequence of this would be to compare the results of the W-B and WAIS to determine what, if any, are the differences. Three such comparisons have appeared in the literature; two utilizing college students and the other psychiatric patients.

Cole and Webela (7) administered both tests to forty-six college students ranging in age from nineteen to twenty-nine years. Although most tests were staggered they report that the WAIS was more often administered as the second test. The IQ's ranged from 105 to 143. Mean IQ's for the first administration were: Verbal Scale 125, Performance Scale 123, and Full Scale 127; mean IQ's for the second administration were: Verbal Scale 127, Performance Scale 130, and Full Scale 130. The greater practice effect was noted on the mean Performance IQ of the second administration. Practice effects, significant beyond the one per cent level, however, were evidenced on all three IQ's of the second test. Practice effects were notably evident on the Picture Arrangement, Object Assembly, and Digit Symbol subtests. One would assume from these findings that the WAIS or W-B should not be used as a retest when either scale has been administered first.

The authors report that the weighted scores on Comprehension and Vocabulary, independent of practice effects, were significantly higher on the WAIS. They presume that this was true of the Vocabulary subtest because the words at the higher end of the WAIS list appeared consistently more familiar to their subjects than those at the higher end of the W-B list. It should be noted, however, that the IQ's of the subjects used in this study appear to tend

modified form of the Matrices Test to eighty-two male patients of a neuro-psychiatric ward who were free of brain-damage. Mean age of the subjects was 31.71; mean years of education was 11.67. Correlation of the modified Matrices with the WAIS Verbal Scale was .684, with the Performance Scale .705, and with Full Scale .721. The higher correlation with the Performance and Full Scales suggests that the modified Matrices Test may be a useful complement to the Verbal Scale in evaluating intellectual functioning of brain-damaged individuals.

Although there have been few research studies published on the WAIS, it appears to this writer, on the basis of those that have appeared, that most research is following the same tendency as was noted in discussing the W-B. This tendency has been to accept the reliability coefficients of the WAIS, as reported by Wechsler, rather than doing reliability studies on comparable normal groups for the purpose of determining whether the reliabilities as reported by him are stable on different population samples. This writer believes it is important to test the stability of the reliabilities reported on all age groups of the standardization study previous to doing research on particular topics using normal or abnormal individuals or groups. The publication of such studies would enable the clinician to have greater confidence in predictions based on interpretation of individual subtest scores and diagnoses based on scatter analysis.

Following publication of the W-B, several studies appeared in the literature relating to short forms of the scale. These short forms usually consisted of different combinations of scores from certain subtests which were used to estimate a Full Scale IQ. For example, one of the most popular short

forms, named "VIBS," combined the scores of the Vocabulary, Information, Block Design, and Similarities subtests of the W-B in order to estimate the Full Scale IQ. A short form of a test, if reliable, is often useful in clinical situations where a quick estimate of intelligence is needed when there is not sufficient time to administer the entire test.

Doppelt (11), one of the psychologists employed in the WAIS standardization, devised a short form of that scale based on information obtained from the records of the 800 subjects of three standardization groups: ages 18-19, 25-34, 45-54. He selected the two best predictors of the total Verbal Score (total of scores on six tests of the Verbal Scale) by correlating all combinations of two Verbal subtests with total Verbal Score. He found the highest correlation between Arithmetic and Vocabulary subtests and total Verbal Score (ages 18-19, .938; ages 25-34, .934; ages 45-54, .948). Following the same procedure with the Performance subtests, the highest correlation was found between Block Design and Picture Arrangement subtests and total Performance Score (ages 18-19, .939; ages 25-34, .917; ages 45-54, .926).

The correlation between sum of scaled scores on these four subtests and the Full Scale Score was then found for the seven age groups comprising the standardization sample and also for the four old-age groups (60-64, 65-69, 70-74, 75 and over) used in standardization of the WAIS on older persons. The coefficients ranged between .95 and .96. The standard deviation of Full Scale Scores was approximately 25; the resulting standard error of estimate was about seven scaled score points, or 4.2 IQ points. Thus, one could expect a Full Scale Score estimated by this method to be within seven scaled points of the actual score about two-thirds of the time.

Regression equations, based on all age groups, were computed for use in predicting Full Scale Score from the four subtest scores. It was found that the predicting variable was similar throughout the age groups and that the constant term of the regression equation varied according to the age of the individual. The predicting variable was set at 2.5 and the constant term ranged from four to ten, depending upon age.

In order to use the short form, an examiner makes use of a simplified regression equation presented by Doppelt (11, p. 65): Estimated Full Scale Score is equal to 2.5 times the sum of scaled scores on the four subtests plus the constant according to age.

Doppelt applied his predictive equations to two groups of subjects not used in the original statistical analysis and found that in 71% of his cases the differences between obtained and estimated Full Scale Scores were within one standard error (±7 scaled points); two standard errors (±14 scaled points) contained 96% of the cases.

This short method appears to obtain a relatively stable estimate of the IQ. Doppelt, however, used a presumably normal population and, in most instances, a short form is most valuable in testing the mentally ill. Olin and Reznikoff (38) and Himelstein (25) have both evaluated the reliability of the short form when used with psychiatric patients.

The first study utilized fifty-four men and forty-nine women patients who had varied schizophrenic and neurotic diagnoses. Mean age of the subjects was 36.5 years. The range of IQ's for the complete WAIS was from 78 to 135, with a mean of 108. Correlation of .925 was found between IQ's obtained from administration of the complete WAIS and IQ's obtained by the short form method.

Standard error of estimate was 7.9 scale points which is comparable to that found by Doppelt.

Himelstein's sample consisted of fifty male patients, thirty-five whites and fifteen negroes, tested upon hospital admission. Mean age of the group was 25.1 years, with a range of twenty-two to sixty-three years. Mean educational level was 8.3 years, with a range of from two to sixteen years. Mean IQ, based on the complete WAIS, was 87.2; mean IQ, based on the short form method, was 85.6. Correlation between IQ's of completed WAIS and those estimated from the short form method was .956; standard error of estimate was 3.5 scale points.

These results from psychiatric samples agree favorably with those obtained from a normal population and suggest that the Doppelt Short Form yields reasonably accurate predictions of IQ.

Two evaluations of the WAIS have thus far appeared in the literature. One by Shafer (46), a clinical psychologist, discusses the test from a clinician's viewpoint. The other reviewer is McNemar (34) whose discussion is from the position of a statistician.

Shafer believes that the outstanding improvement of the WAIS is the "all-new Vocabulary list containing mostly verbs, adjectives, and abstract or 'literary' nouns" (46, p. 157). He regards the W-B Vocabulary as containing many items, such as "guillotine," which overlap Information items, while the WAIS list more directly obtains an idea of the subject's verbal self-expression and defines his verbal organization and communication of experience. On the other hand, he notes that the Information subtest of the WAIS now contains two items overlapping Comprehension ("Why are dark clothes warmer than light-

colored clothes?" and "How does yeast cause dough to rise?"); and the comprehension subtest contains three items (proverbs) which overlap concept formation or the Similarities subtest. Thus, he believes, all changes in the WAIS have not been toward achieving greater homogeneity within subtests.

Also commendable in the new scale, according to this reviewer, is the use of "lead-off" items at the lower end of most subtests. These enable an examiner to obtain a better estimate of intellectual functioning at the lower levels. Another improvement is the comparing of an individual's test performance with the average of his peers, rather than with the average of a standard reference group. This supplants the Efficiency Quotient of the W-B whereby a testee's performance was compared with the average of the 20-34 age group to determine efficiency of intellectual functioning.

This reviewer comments that the lengthening of most WAIS subtests is disadvantageous in that more time is required to administer the complete scale. Consequently, there will be a tendency on the part of some clinicians to use a short form. On the other hand, greater length is advantageous in that a broader sample of behavior is obtained.

Shafer is critical of the fact that the Digit Span and Arithmetic subtests remain a part of the Verbal Scale. It has been his experience that both subtests are very vulnerable to psychopathology. Often, he states, a much better clinical estimate of the patient's verbal level can be obtained by omitting these two scores and intrapolating. He regrets that the new scale does not include a subtest for the purpose of measuring mental deterioration, such as a measure of immediate memory for meaningful material. He regards neither the Digit Span nor Digit Symbol subtests as being satisfactory in this regard.

As a clinician attempting to use the WAIS to understand an individual, Shafer is critical of Wechsler's goal to eliminate certain ambiguous items from the scale. He states that those W-B items eliciting frequent error, such as "How many pints in a quart?," often added to the understanding of a patient's temporary inefficiencies, confusional tendencies, and uncertainty. One can presume, however, that similar ambiguous items will be found in the new scale.

Shafer does not believe that it will be difficult for clinicians to switch from the W-B to the WAIS because, even though there have been changes, the type of items remains the same.[4] He concludes by stating that a definitive clinical assessment must await considerable research with a variety of clinical and normal subjects.

McNemar (34), a highly regarded statistician, evaluates the standardization sample of the WAIS as being excellent and believes it a marked improvement over that of the W-B and other scales of the same vintage. He accepts the split-half reliabilities presented by Wechsler of the eleven subtests but is somewhat skeptical of the reliability coefficients for Verbal (.96), Performance (.93), and Full Scale IQ's (.97). The prediction is made that test-retest reliabilities will not be as high on the IQ scales.

McNemar chides the author of the WAIS for showing no recognition of the vast factor-analysis literature which, if anything, has consistently indicated that a "test constructor should strive for a pure measure of whatever he hopes to quantify. That is, a score should represent a point on a unidimensional scale rather than a hodgepodge of different dimensions" (34, p. 159).

In the preceding evaluation, Shafer, a clinician, criticized the inclusion of the Digit Span and Arithmetic subtests in the Verbal Scale of the WAIS

because these tests were vulnerable to distortion with psychiatric patients. This reviewer, a statistician, makes the same criticism but on the basis that these two subtests are not true measures of verbal ability or, in other words, they are not pure measures on a unidemensional scale.

McNemar believes that Wechsler undertook an impossible task in attempting to construct a scale to measure general or, as Wechsler named it, global intelligence which will also provide differences among subtests that will be of diagnostic value. He states that the diversity of content within the eleven subtests is too great and that the intercorrelations between them are too low to satisfy the requirements for a reliable measure of general intelligence. On the other hand, the diversity of the subtests is not great enough, in the factor-analysis sense, to yield the low intercorrelations necessary for reliable difference scores that will be of diagnostic value.

The research and evaluations of the WAIS that have been presented reflect critical as well as promising aspects of the new scale. It cannot be doubted that the W-B, in spite of its limitations, was an advancement in testing adult intelligence. One can be certain that the WAIS will also be found to have limitations. Intelligence scales, however, are a necessary and valuable complement to the function of the clinical psychologist. Research with the WAIS should endeavor to discover the limitations as well as the advantages in order that the working clinician can better evaluate his subjects. It is the purpose of this study to provide, in part, some of this needed information.

D. Comparison of the WAIS with the SBL.

It was stated previously that a recent survey showed the SBL to be the most used intelligence scale in the United States. This survey, however,

obtained more information from child guidance than adult centers and most psychologists definitely would prefer the W-B or WAIS when testing adults. There are instances, however, when knowledge regarding the comparability of these adult scales and the SBL is valuable. For example, when retesting is necessary and the W-B has been the first test it is impractical to use that scale as the retesting instrument because of practice effects such as higher scores on timed items and on problems involving simple learning. One would expect similar practice effects if the WAIS were used as a second test because of the similarity of the two scales. Possibly the SBL could be used in retesting. Another instance where information regarding comparability would be important, would be when a comparison is made between the adult scales and the SBL. For example, if a mental patient had taken the SBL as a child and as an adult was administered one of the adult scales, the results of the two scales might reveal what amount of intellectual deterioration, if any, has occurred. It is advantageous then to have some understanding of the comparability of the WAIS and SBL.

At the present time, one WAIS-SBL comparison has been reported. This was done by Wechsler (51) on a randomly selected group of fifty-two white males who were reformatory residents. Age of subjects ranged from sixteen to twenty-six years of age; mean years of education was 8.3 years. The tests were administered by two examiners according to a plan so that essentially the same interval of time occurred between tests and so that each examiner gave an equal number of WAIS and SBL tests.

Mean IQ's for the SBL, Verbal, Performance, and Full Scale were 100.5, 93.8, 98.3, and 95.4, respectively. Standard Deviation for the SBL was 10.9;

for the Verbal Scale, 10.9; for the Performance Scale, 12.6; and for Full Scale, 11.7. Correlation of the SBL with the three WAIS scales was as follows: Verbal .86, Performance .69, Full Scale .85.

The high correlations found between the SBL IQ's and the Verbal and Full Scale IQ's of the WAIS conform to expectation. The WAIS Performance Scale IQ, as expected, correlates to a lesser extent with the SBL IQ. Similar results have been found in studies correlating the SBL IQ with the Performance Scale IQ of the W-B (2, 48). This finding appears to be due to the high verbal content of the SBL.

The difference between mean IQ's of these two tests is interesting because the SBL mean IQ is about equal to that of the general population, whereas the mean Full Scale WAIS IQ is below the general population average. The standard deviation of 17.3 on the SBL is also close to that of the general population standard deviation of about 16 on the SBL. The standard deviation of 11.3 on the WAIS Full Scale IQ, however, is considerably lower than the general population standard deviation of 15 on the WAIS.

Wechsler makes the assumption that the WAIS mean IQ's are a better representation of this group than the SBL mean IQ. This assumption is based on the fact that the educational level of the group is inferior to that of the general population average. This inferiority would be expected to be reflected by the resulting mean IQ's of each test because both the WAIS and SBL correlate highly with educational level. The SBL mean IQ for this group does not, however reflect educational inferiority. The WAIS IQ's do reflect the expected restriction in variability of IQ and Wechsler thus suggests that the WAIS IQ's give a better representation of the group.

Whether this is true or not depends on the results of further research wherein WAIS IQ's are compared with IQ's obtained from the SBL and other tests. It may be that the disparity in mean IQ's, such as seen in Wechsler's study, are due to real differences between tests rather than one test actually presenting a better picture of the group. Furthermore, the group used by Wechsler for this study cannot be assumed to be comparable to a randomly selected population sample and could be responsible for the noted differences.

Although there has been no other published comparison of SBL and WAIS results, there has been considerable research on the correlation between W-B and SBL IQ's. One would expect these studies to reflect what might be expected of WAIS-SBL comparisons, since the WAIS is considered to be a modification of the W-B, and following are the results of some of these studies.

Weider, et. al, (52) reports a correlation of .81 between the Full Scale IQ of the W-B and SBL IQ. The population was sixty-one delinquent white males, eight to sixteen years of age. Mean age was twelve years. Correlation of the Verbal Scale IQ of the W-B with the SBL IQ (.87) was higher than the Performance Scale IQ (.56). Correlation between Full Scale and SBL IQ's was .81. The authors state that most of the brighter individuals obtained lower IQ's on the W-B than they did on the SBL, while the duller subjects obtained higher IQ's on the W-B than they did on the SBL, while the duller subjects obtained higher IQ's on the W-B. Most investigators (14, 16, 45, 47) who have correlated the SBL IQ with the Verbal, Performance, and Full Scale IQ's have found that brighter subjects usually do better on the SBL and that duller subjects obtain higher IQ's on the W-B.

Goldfarb (14) administered the SBL and W-B to sixty foster children,

ranging in age from eleven to seventeen years; mean age was fourteen years. Correlation for Full Scale IQ with the SBL IQ was .86; Verbal IQ with SBL IQ was .80; Performance IQ with SBL IQ was .67. He found that the superior adolescents averaged five points higher on the SBL than on the W-B.

Sartain (45) administered the two tests to fifty-one college freshmen. The age range was not reported. Correlation coefficients between the SBL IQ and Full Scale IQ was .77; between SBL IQ and Verbal IQ .80; and between SBL IQ and Performance IQ .51. He, as did Goldfarb (14), reported that with the superior students the SBL IQ's were significantly higher than those obtained on the W-B.

Guertin (16), found the W-B Full scale to give scores approximately five IQ points higher than the SBL. His eighty-two subjects were from fifteen to twenty-two years of age and ranged from high-grade defective to dull normal in intelligence. This finding is also that of Stacy and Levin (47) from tests administered to 100 retarded individuals ranging from ten to twenty-three years of age.

Kutash (30) tested fifty adult institutionalized male mental defectives. The subjects ranged in age from sixteen to fifty-nine years of age; mean age was 36. He found the following correlations with the SBL IQ: .76, .73, .51 for Full Scale, Verbal, and Performance IQ's respectively.

The majority of studies report that the brighter subjects obtain lower IQ's on the W-B than on the SBL. The duller subjects, however, obtain higher IQ's on the W-B than on the SBL. One reason for this is that the SBL is generally considered to be a test in which academic learning, or verbal skill, is important. It would be expected then, that the brighter subjects, who have

probably attained a higher educational level than the duller ones, would tend to do better on the SBL. The duller subjects, on the other hand, have a greater opportunity to obtain a higher IQ on a scale where academic learning is not important throughout the test. The W-B IQ is obtained from Verbal and Performance scales and the Performance Scale, wherein success does not depend as greatly on academic learning as it does on the Verbal Scale, permits the duller subject to obtain a higher score.

It will be noted that the correlations between SBL and Full Scale IQ's are lower than expected in those studies where brighter or duller subjects comprised the sample. This would be expected because brighter subjects do better on the SBL and this would have the effect of lowering the correlation between tests. Similarly, in studies utilizing duller subjects, higher scores would be obtained on the W-B, and again the correlation would tend to be lowered. In those studies employing a presumably normative group, however, correlations between Full Scale and SBL IQ's are high and relatively stable from study to study. Correlation between Performance Scale and the SBL are consistently low. This fact is generally considered to be the result of the verbal content of the SBL in contrast with the non-verbal Performance Scale.

In the following studies hospital patients were used as subjects. Here also, in spite of possible distortion because of mental illness, correlations between the two tests are high.

Halpern (22) tested 133 clinic out-patients ranging in age from ten to past thirty-five. She divided the subjects into four age groups. Group A numbered thirty-seven children, age ten to fourteen. Correlation between the SBL IQ and Full Scale IQ was .92. Group B consisted of fifty-five persons, age fifteen to

twenty-four. Correlation between SBL IQ and Full Scale IQ was .91. Group C numbered twenty-three persons, age twenty-five to thirty-four. Correlation was .90. Group D consisted of those persons thirty-five years of age and over. There were eighteen persons in this group and correlation was .90.

Benton, et. al. (4), administered the SBL and W-B to sixty mental hospital patients ranging in age from sixteen to fifty-nine years; mean age was 35. Correlations between SBL IQ and Verbal, Performance, and Full Scale IQ's were .92, .73, and .93, respectively.

Mitchell (36), administered the tests to 227 state hospital psychotics and alcholics ranging from fifteen to sixty-five years of age. Mean age was approximately thirty-seven years. The SBL IQ correlated with Verbal Scale IQ was .91; with Performance Scale IQ was .80; with Full Scale IQ was .89.

There is considerable dispersion in regard to age, range of intelligence, and normality in the several studies that have been reviewed. Most findings reflect high comparability between the SBL and W-B Verbal and Full Scale IQ's. In view of the similarity between the WAIS and W-B, a comparison between the WAIS and SBL IQ's would be expected to reflect corresponding results. Likewise because of the highly verbal content of the SBL, one would expect a lower correlation between SBL and Performance Scale IQ's. It will be interesting to discover whether the changes that have been made in the WAIS will result in different findings for the normative group used in the present study or whether they will conform to previous correlations between the W-B and SBL.

CHAPTER III

DESIGN OF RESEARCH

The WAIS and SBL were administered to one hundred subjects, fifty males and fifty females. Each person selected had passed his eighteenth birthday but had not reached his twentieth birthday. All except fourteen of the testings were completed by the writer. The two examiners who administered the fourteen testings were experienced in administering both the WAIS and the SBL. All testings were done in a professional setting. Both tests were administered the same day and the order of administration was staggered.

A. Selection and Procurement of Sample

1. Criteria for Selection of Subjects.

The subjects were selected according to the criteria presented below. The percentages following certain of the categories indicate the percentage of males and females included in each category. These percentages were computed according to the 1950 United States Census with special consideration given to the census figures for the city of Chicago.

a. Age: 18.00 to 19.99 years of age.

b. Sex: 50 males, 50 females.

c. Race: 86% white, 14% non-white.

d. Occupation:

(1) Professional, technical and kindred workers. (1% male, 2% female)

(2) Managers, officials, and proprietors, except farm. (2% male, 0% female)

(3) Clerical, sales, and kindred workers. (6% male, 13% female)

(4) Craftsmen, foremen, and kindred workers. (4% male, 0% female)

(5) Operatives and kindred workers. (8% male, 3% female)

(6) Private household workers. (0% male, 2% female)

(7) Service workers, except private household. (2% male, 5% female)

(8) Laborers except farm. (11% male, 0% female)

(9) Keeping house. (0% male, 15% female)

(10) Students. (14% male, 11% female)

(11) Others. (working less than 15 hours weekly, unable to work, voluntarily idle.) (2% male, 1% female)

e. Education:

(1) Completed eight years of education or less. (12% male, 8% female)

(2) Completed nine to eleven years of education. (17% male, 18% female)

(3) Completed twelve years of education. (16% male, 21% female)

(4) Completed thirteen or more years of education. (4% male, 4% female)

f. Religion (based on information obtained from the Chancery Office, Archdiocese of Chicago):

(1) Protestant 54%

(2) Catholic 40%

(3) Jewish 6%

(4) Other 1%

2. Procedures in Obtaining Sample.

None of the subjects were known personally by the examiners. The majority were obtained by contacting persons employed in personnel departments of various businesses, school principals and teachers, directors of youth programs, and union officials. Some subjects were directly contacted, i.e., gasoline station attendants. Another source was the testees themselves. In a few instances they told the tester of friends who would be willing to participate.

The usual procedure of contacting the subjects was as follows. The writer contacted the sources mentioned above and explained the salient features of the study. They were asked to supply names and telephone numbers of possible subjects. Upon obtaining this information, the writer would telephone the prospect, identify himself, and state the source from whom the prospect's name was obtained. A brief explanation of the study was then given. This usually consisted of stating that the writer was making a comparison of the responses of eighteen and nineteen year old individuals to an old and a new test in order to determine how accurately the new test measured. The prospective subjects were informed that this was a scientific study and the names of the persons participating would not be used. It was further explained that completion of the two tests would require between one hour and a half to

two hours.

The only instance where greater explanation was required, and this was concerning the identity of the writer and his purpose, occurred when the prospective testee was female and a parent, usually the mother, asked for elaboration.

The prospects were not told that the tests were measures of intelligence. A few subjects inquired as to the type of questions that would be asked and this was handled by stating that the questions would not be of a personal nature. This explanation apparently was satisfactory for they did not pursue their inquiry. During the actual testing, some subjects indicated that they knew an intelligence test was being administered. This was neither affirmed nor denied; but if the subject specifically asked if he were being given an intelligence test, he was told that he was.

If the individual agreed to participate he was given a choice of possible appointment times. The writer's employment enabled him to test subjects during the working day, in the evening, or on Saturday. After an agreeable time had been arranged the location of testing was decided and an explanation was given regarding the best method of reaching the location. The testing was done in one of the laboratory booths of the Psychology Department of Loyola University or in the offices of the Guidance Department of Catholic Charities.

A postcard was sent to the subject, following the scheduling of an appointment, so that it would be received the day previous to the appointment. This was done to reduce appointment failures. The exact effect of receiving the postcard is not known. Several subjects commented, however, that they would have forgotten had it not been for the postcard. The content of the

postcard was as follows:

"Dear _____:

According to our telephone conversation an appointment has been scheduled for you at _____ o'clock, on (day), (month) (date), at (location), (address).

If you are unable to keep this appointment please call CEntral 6-5172 and cancel it. If I am not available, please tell the secretary and she will inform me. I will call you so that another appointment can be made. Thank you.

(signed) Fred D. Whelan"

3. Response of Persons Contacted

A total of 142 contacts were made. Of this number, twenty-eight persons, twelve males and sixteen females, stated they were not interested in participating. Eleven of the remaining 114, five males and six females, agreed to participate but did not appear at the designated time. Follow-up telephone calls were made to all persons who did not appear for the first appointment. In the case of these eleven persons, however, they did not appear even when a second or third appointment was scheduled. In the case of other subjects who failed the first appointment, a second appointment was usually kept. These latter subjects gave two reasons for failure: 1) they had either forgotten the appointment or 2) a situation more important to them prevented their appearance.

Of the remaining 103 testees, two female subjects appeared for their appointment but decided that they did not wish to continue shortly after the

testing had begun. The writer used every possible technique to obtain cooperation from all subjects. These two, however, were adamant in their refusal to complete the tests. Of the remaining 101 subjects, one female subject came for her appointment but became ill after one test had been completed and the second one begun. She telephoned a week later and stated that she was to be hospitalized and could not complete the second test. The number of individuals not tested and reasons for their non-participation are presented in Table 4.

Table 4

NUMBER OF INDIVIDUALS NOT TESTED, CLASSIFIED BY REASON AND SEX

Reason	Male	Female	Total
Refusing to participate	12	16	28
Not appearing for appointment	5	6	11
Refusing to complete tests		2	2
Unable to complete tests		1	1
Individuals not tested, both sexes			42

It required approximately nine months to complete the testing of the one hundred subjects. The first tests were administered in May, 1956, and the last in January, 1957.

4. Difficulties Encountered in Obtaining Subjects.

There were two difficulties encountered in obtaining subjects to fulfill the criteria of population selection. The first difficulty was obtaining male

subjects who could be classified as "Laborers" and the second was obtaining females who could be classified as "Keeping House."

The first problem was eliminated by using males who were just beginning their apprentice training and who, according to the estimates of their employers, were not likely to complete or continue their training. Employer evaluation of these subjects reflected poor educational preparation, lack of necessary skill to succeed, and poor initiative or motivation. In the opinion of their employers, these men would probably continue to be employed by them, but as laborers rather than apprentices.

Population criteria required fifteen female subjects who were "Keeping House." It was possible to obtain ten females whose principal occupation was caring for their home. The five additional subjects used to fulfill the requirements of this category were females who worked part-time and also cared for their home. The reason for working in the case of these five subjects was financial necessity rather than not wanting to be a homemaker. All fifteen subjects in this category were married.

5. Scoring of the Tests Administered.

The accuracy of scoring on each test was reviewed by three persons. First, the person administering the test reviewed his scoring. The tests were then reviewed by two other persons experienced in administering and scoring the WAIS and SBL. Careless errors, such as incorrect counting, were handled directly by the reviewer. However, when a reviewer disagreed with the score given a particular item that was scored on the basis of what the testee said, i.e., a response to a comprehension item on the Verbal Scale of the WAIS, a note was made of the disagreement. After all three persons had reviewed the tests, the

disagreements in scoring were discussed and resolved. The scoring concepts outlined in the WAIS and the SBL manuals were strictly adhered to in the resolution of these disagreements.

B. Statistical Analysis of the WAIS and SBL

1. Intra-reliability of the WAIS.

a. The reliability of each subtest, except Digit Span and Digit Symbol, was computed by determining the correlation by the Pearson Product-Moment Method between scores on odd and even items and correcting the coefficient for full length of the test by the Spearman-Brown formula.

b. The reliability of the Digit Span subtest was computed by two methods. The first method, used by Wechsler, was to obtain the correlation between Digits Forward and Digits Backward and to correct this correction for full length of the test by the Spearman-Brown formula. The second method was determined by the writer because it was believed that Wechsler's method of correlating Digits Forward and Digits Backward carried the assumption that the same ability was equally measured by both parts of the subtest. This might be considered true if the ratio was constant for most persons between the number of digits successfully repeated forward and the number of digits successfully repeated backward. This ratio, however, is not constant. It would appear that a great deal more concentration is required to successfully repeat digits backward. The individual must not only be able to retain the digits as given, but must retain them in a correct order forward until he can successfully recall them in the reverse order.

The writer, considering that greater weight should be given to the successful repetition of digits backward, devised the following method of

estimating the reliability of the Digit Span subtest. The number of digits successfully repeated forward was assigned odd numbers. Thus repeating three digits forward was assigned a value of one; repeating four digits forward was given a value of three, etc. It was assumed that repeating two digits backward was as difficult as repeating three digits forward. Thus, two digits successfully repeated backward was assigned a value of two; three digits backward was given a value of four, etc. It is believed that values assigned in this manner permit more equitable weighing of the number of digits successfully repeated backward. Below is an example of the Digit Span subtest of the WAIS with the numerical values assigned to successful repetition of digits forward and backward.

Digit Span Subtest - WAIS

Digits Forward	Score	Digits Backward	Score
582	1	24	2
6439	3	629	4
42751	5	3279	6
619473	7	15286	8
5917428	9	539418	10
58192747	11	8129365	12
275862584	13	94376258	14

Thus, if a subject correctly repeated five digits forward, his score would be 5; if he successfully repeated five digits backward, his score would be 8. The reliability of this subtest was determined by correlating the odd and even

scores and correcting the obtained correlation for full length of the test by the Spearman-Brown formula.

c. The method used by Wechsler in determining reliability of the Digit Symbol subtest has been described in Chapter II, "Related Literature." The manner in which this reliability coefficient was established leads one to believe that Wechsler considered the Digit Symbol subtest of Form II of the Wechsler-Bellevue Intelligence Scale to be equivalent to the WAIS Digit Symbol subtest. How this assumption was obtained is not explained by Wechsler in his presentation of the standardization data. Nor does he explain why the Digit Symbol subtests were administered to female rather than to both male and female subjects.

In the present study, rather than assuming that the WAIS Digit Symbol subtest has an equivalent form, the writer used a different technique for estimating the reliability of this subtest.

One of the feasible methods of determining reliability of a speed test is to divide the total time into quarters and to find a score for each of the four quarters. The scores of the first and fourth quarter are then summed to yield one half-score or odd score and the scores of the second and third quarters are summed to represent a second half-score or even score. The odd and even scores are then correlated and the obtained coefficient corrected for full length of the test.

The time limit of the WAIS Digit Symbol subtest is 90 seconds. This time limit, divided in quarters, yielded four time intervals, each being 22.5 seconds in length. While the test was being administered the examiner closely observed the performance of each testee and noted the number of items completed within

each time interval of 22.5 seconds. The number of items completed in the first and fourth quarters were combined to yield one-half score, and the items completed in the second and third quarters were combined to represent the other half-score. The obtained Product-Moment correlation coefficient between these two half-scores was corrected for full length of the test by the Spearman-Brown formula in order to estimate reliability of this subtest.

It was assumed that the combination of time quarters tended to balance out the cumulative effects of practice, fatigue, and other factors. One problem of timing each quarter was the scoring of symbols that were begun but not completed within the time interval of each quarter. This problem was met by always crediting a symbol to the time interval in which it is begun. It was assumed that any differences in half-scores resulting from such a method of scoring would tend to be balanced out in the administration and scoring of one hundred tests.

d. The reliability coefficients of the Verbal, Performance, and Full Scale IQ's were estimated by two methods in this study. The first method was that utilized in the standardization study. The second method was that employed by the writer because certain assumptions made in obtaining the coefficients in the standardization study were not considered acceptable.

Wechsler used a simplified version of the formula for correlation between sums of equally weighted scores to obtain the reliability coefficients of the Verbal, Performance, and Full Scale IQ's (28, p. 396).[4] This formula was

---

[4]Doppelt, J.E. Personal Communication, 1957.

simplified by making certain assumptions. These were: that the standard deviations of all subtests were equal and that the intercorrelations between all subtests were equal. It was possible, on the basis of these assumptions, to exclude these terms from the basic formula and the following simplified version of the formula was the result:

Correlation between odd-half and even-half tests:

$$r = \frac{\sum_{1}^{A} r_{ij} + \sum_{1}^{A^2-A} r_{ab}}{A + \sum_{1}^{A^2-A} r_{ab}}$$

where $\sum_{1}^{A^2-A} r_{ab} = \frac{1}{2} \sum_{1}^{A^2-A} r_{xy} \sqrt{1 + r_{xx'}} \sqrt{1 + r_{yy'}}$

A = number of tests

$\sum r_{ij}$ = correlation between odd and even items for each test; also appears as $r_{xx'}$ and $r_{yy'}$

The obtained coefficient is corrected for full length by the Spearman-Brown formula.

The assumptions made in order to obtain this simplified formula were not considered acceptable because the standard deviations of the subtests and the intercorrelations between subtests are not equal (See Appendix, Tables 10 through 19). Thus, the basic formula for correlation between sums of equally weighted scores (28, p. 396) was the second method employed in this study and no assumptions of equality were made between the standard deviations of all subtests and intercorrelations of all subtests. The obtained coefficient is corrected for full length by the Spearman-Brown formula.

It will be interesting to compare the reliability coefficients obtained through these two methods to determine whether there is a significant difference between them.

2. Correlation of the WAIS with SBL.

Correlation coefficients between the SBL and the Verbal, Performance, and Full Scale IQ's will be determined by the Pearson Product-Moment method.

CHAPTER IV

ANALYSIS OF RESULTS AND INTERPRETATION

A. Representativeness of the Sample.

The WAIS and SBL were administered to one hundred individuals according to the stratification plan outlined in Chapter III, "Design of Research." This plan, based on an analysis of figures from the 1950 United States Census, was followed in an effort to obtain a representative sample of the population.

The fifty males and fifty females tested were between eighteen and nineteen years of age. Mean age was 18 years, 8 months; standard deviation was 7.08. Of the one hundred subjects, eighty-six were white and fourteen were non-white. Fifty-two subjects were of Protestant faith, forty were Roman Catholic, and six were Jewish. Two of the subjects stated that they did not belong to a particular religious group.

The percentage of subjects included in each occupational category is presented in Table 6. The percentage of individuals within each category according to the 1950 census is also presented in the same table. The percentages shown for each occupational group indicates close agreement between the national percentage and the percentage included in this study.

The percentage of subjects within the four educational levels represented in this study are presented in Table 6. Population percentages according to the 1950 census are also presented in the same table. The greatest difference in population and sample percentages occurred in the 9-11 and 12 years levels for

Table 5

PER CENT OF SUBJECTS ACCORDING TO OCCUPATIONAL CATEGORY

Occupational Category	Male N=50		Female N=50	
	Percentage in U.S. Pop.	Percentage in Sample	Percentage in U.S. Pop.	Percentage in Sample
Professional, technical, kindred workers	1	2	3	4
Managers, proprietors, except farm	3	4	*	0
Clerical, sales, and kindred workers	11	12	25	26
Craftsmen, foremen, and kindred workers	6	8	*	0
Operatives and kindred workers	17	16	7	6
Private household workers	*	0	3	4
Service workers, except private household	4	4	5	6
Laborers, except farm	24	22	1	0
Keeping House	0	0	30	30
Students	28	28	25	22
Others - unable to work, voluntarily idle	5	4	2	2

*Less than 1%

the males and females. There is a greater percentage of males in the 9-11 years level while there is a lesser percentage in the 12 years level. The females on the other hand, have a greater percentage in the 12 years level and a lesser percentage in the 9-11 years level. Mean years of education for males was 10.54, standard deviation was 1.96; for females 11.32, standard deviation was 1.58.

Table 6

PER CENT OF SUBJECTS ACCORDING TO EDUCATIONAL LEVEL

Years of School Completed	Male N=50		Female N=50	
	Percentage in U.S. Pop.	Percentage in Sample	Percentage in U.S. Pop.	Percentage in Sample
8 or less years	25	26	18	14
9 - 11 years	34	38	32	30
12 years	33	28	41	48
13-15 years	8	8	9	8

Examination of the number and percentage of subjects within the six criteria comprising the stratified sampling plan indicates that except for the difference between the educational level of male and female subjects, the individuals employed in this study appear to be a representative sampling of the national 18-19 year age group.

B. Reasons for using the Split-half method.

There are several methods of estimating reliability coefficients, such as the retest, comparable forms, and split-half methods. There is no one best way of estimating reliability. The method employed depends upon the purpose and use of the coefficients and also the practical considerations of the research situation.

In this study the comparable form method could not be utilized because there was no comparable form of the WAIS. The retest method has several disadvantages. For example, the administration of an individual intelligence test to one hundred randomly selected subjects requires considerable time. The probability of being able to retest these same individuals after a lapse of time would be doubtful. In retesting there is also the problem of practice effects which are often statistically significant. These effects lessen in time but it is difficult to ascertain when the practice effect will no longer distort reliability coefficients. Also to be considered in using the retest method are the changes in examiner and examinee, in the interval of time between tests, which may affect reliability coefficients.

The split-half method was used in this study for several reasons. First, it can be applied to a single administration of a test. Secondly, there is assumed to be less error variance in using this method because changes in conditions during a single test administration are considered to be uniformly distributed over all items in both halves of the test. Third, there are no practice effects. Fourth, since individual intelligence tests are usually administered at one sitting, it would seem reasonable to assume that a reliability coefficient gained by the split-half method, which utilizes a single

administration, would be closer to the basic concept of measuring reliability.

C. Reliability of WAIS subtests and IQ's.

The means, standard deviations, reliability coefficients, and standard errors of the reliability coefficients of WAIS subtests and Verbal, Performance, and Full Scale IQ's obtained on the one hundred subjects used in this study are presented in Table 7. The reliability coefficients obtained on the standardization group are also presented in the same table. Means and standard deviations were not reported for the standardization group.

The reliability coefficients of each subtest and IQ scale obtained for the experimental and standardization group were subjected to the z test of differences between correlations (18, p. 224). None of the differences between reliability coefficients for each subtest and IQ scale were found to be significant at the five per cent level. The standard errors of the reliability coefficients (18, p. 205) indicate that none of the obtained reliability coefficients deviate significantly from the true population value at the five per cent level. Since the results of the present study conform to those reported on the standardization group, it can be assumed that the reliability of WAIS subtests and IQ scales on an 18-19 year age group as published by Wechsler are stable.

It was explained previously in Chapter III, "Design of Research," pages 41, 42, and 43, that the reliability coefficient for the Digit Span subtest would be estimated by two methods. One method, that used by Wechsler, was to correlate digits forward with digits backward. The second method, that devised by the writer, was to assign numerical values to the number of digits correctly repeated forward and backward. These values, representing odd (Digits Forward)

Table 7

MEANS, STANDARD DEVIATIONS AND RELIABILITY COEFFICIENTS OF WAIS

Tests	Experimental Group N=100				Standardization Group N=200
	Mean	SD	r	σ_r	r
Information	14.20	2.80	.88	.022	.91
Comprehension	18.10	2.69	.79	.038	.79
Arithmetic	10.54	1.84	.82	.044	.79
Similarities	13.92	2.59	.83	.032	.87
Digit Span	14.63	2.48	.71	.050	.71
Vocabulary	41.22	8.48	.95	.009	.94
VERBAL SCALE IQ	103.44	14.56	.96	.008	.96
Digit Symbol	55.91	6.28	.95	.009	.92
Picture Completion	11.80	2.22	.76	.049	.82
Block Design	32.05	5.51	.87	.024	.86
Picture Arrangement	23.50	3.67	.66	.057	.66
Object Assembly	28.45	5.16	.71	.049	.65
PERFORMANCE SCALE IQ	99.45	14.21	.93	.013	.93
FULL SCALE IQ	101.90	15.33	.97	.007	.97

and even (Digits Backward) scores were correlated and the coefficient corrected for full length. This latter method was devised in an effort to provide more equitable weighting to the more difficult task of repeating Digits backward.

The reliability coefficient of the Digit Span subtest reported in Table 7 is that obtained by the second method. The coefficient obtained by correlating

Digits Forward with Digits Backward was .735 with a mean of 10.42 and a standard deviation of 7.79. The difference between the coefficients (.735, .706) obtained by the two methods was not significant. This attempt to provide more equitable weighting of digits backward in relation to digits forward did not significantly affect the reliability of this subtest.

It was stated in Chapter III, "Design of Research," that Wechsler had used a simplified version of the formula for the correlation between sums (28, p. 396) to obtain the reliability coefficients of the three IQ scales of the WAIS. In order to use the simplified formula, Wechsler assumed that the standard deviations of each subtest and the intercorrelations between subtests were equal. Since these assumptions were not true, it was believed that the high reliability coefficients reported in Table 7 for the three IQ scales on the standardization group were a result of making these assumptions. Thus, in this study, it was decided to determine the reliability coefficients by two methods: using the simplified formula applied by Wechsler and using the original formula.

The reliability coefficients for the three IQ scales presented in Table 7 are those obtained by using the original formula (Appendix, Tables 10 through 19). The reliability coefficients obtained for the three IQ scales when the simplified formula was used are as follows: Verbal Scale IQ .960, Performance Scale IQ .872, and Full Scale IQ .972. The z test of differences was applied and it was found that there was not a significant difference at the five per cent level between the reliability coefficients obtained by using the original and those obtained by using the simplified formula.

It can be assumed then, that the reliability coefficients published by Wechsler are stable estimates of the reliability of the WAIS Verbal, Performance,

and Full Scale IQ's of an 18-19 year age group.

D. Correlation between the WAIS and the SBL.

The IQ's of the one hundred subjects obtained from the three IQ scales of the WAIS and the SBL were correlated by the Pearson Product-Moment method. The coefficient of correlation between the Verbal Scale and SBL was .921; between the Performance Scale and SBL was .811; between the Full Scale and SBL was .925. Standard error of the difference of the correlation coefficients between Verbal Scale IQ and SBL IQ was .015; between Performance Scale IQ and SBL IQ was .034; between Full Scale IQ and SBL IQ was .015. These values indicate that none of the obtained correlation coefficients deviate significantly from the true population value at the five per cent level.

As would be expected, there is a high correlation between the SBL IQ and the Verbal and Full Scale IQ's of the WAIS. The Performance Scale IQ correlates to a lesser extent with the SBL IQ. These correlation coefficients are higher than those reported in similar studies correlating SBL and W-B IQ's. This is especially true of the correlations between SBL and W-B Performance Scale IQ's. The correlation coefficient of .811 reported in this study between SBL and WAIS Performance Scale IQ's, in contrast with previously reported lower coefficients, would seem to indicate that changes in the WAIS have increased the verbal content of the Performance Scale. If this correlation of .811 is comparable with those reported in similar future studies, the measurement of performance abilities as being distinct from the measurement of verbal abilities has been seriously weakened by changes in the WAIS.

Wechsler (51, p. 21) reports a coefficient of .69 between SBL and WAIS Performance Scale IQ's. One can assume that his restricted population probably

accounts for this low correlation. The fact that the population sample used in the present study closely conforms to the national population in sampling, mean IQ, and standard deviation would cause one to believe that the coefficient of .811 found in this study between SBL and WAIS Performance Scale IQ's would be closer to the true relationship between the two scales.

The mean IQ's and standard deviations of the WAIS and SBL are presented in Table 8. The mean IQ's and standard deviations of the Performance and Full Scales of the WAIS are close to the general population average of 100 and standard deviation of fifteen. Although the mean Verbal Scale IQ is above that of the general population mean, the standard deviation is approximately that of the general population standard deviation. A t-test of differences (18, p. 228) indicated that there was not a significant difference between the mean IQ's obtained in this study and those published by Wechsler for the Verbal, Performance, and Full Scale IQ's. An F-test of differences (18, p. 232) indicated that there was not a significant difference between the standard deviations obtained in this study and those published by Wechsler for the Verbal, Performance, and Full Scale IQ's.

The mean of the SBL IQ is above that of the general population mean. A t-test of differences indicated that there was not a significant difference between the obtained SBL mean IQ and that of the population mean. The standard deviation of the SBL is also above that of the general population standard deviation of sixteen. An F-test of differences indicated that the difference between the standard deviation obtained in this study and that of the general population was significant at the one per cent level.

The difference in variability of the SBL IQ's, in contrast with those of

the WAIS, is probably because the WAIS is a better measuring instrument when both tests are administered to an eighteen-nineteen year old sample. Previous research, Chapter II, pages 32 and 33 has shown that brighter subjects tend to obtain higher IQ's on the SBL and duller subjects tend to obtain lower IQ's on the SBL. The larger percentage of subjects in the duller and brighter IQ groups would tend to increase the variability of the SBL when administered to a sample such as was used in this study. Another possible factor may be that the SBL standard deviation of sixteen was obtained from a standardization sample consisting primarily of children. In other words the sample tested in the present study is not comparable to the sample upon which the SBL was standardized and thus different variability in IQ's could be expected. The WAIS, being standardized on the general adult population, would be expected to agree with the results of the present study.

Table 8

MEAN IQ'S AND STANDARD DEVIATIONS OF WAIS AND SBL

TEST	Mean IQ	SD
WAIS		
VERBAL SCALE	103.44	14.56
PERFORMANCE SCALE	99.45	14.21
FULL SCALE	101.90	15.33
STANFORD-BINET	104.00	20.79

One can assume that there is comparability in mean IQ's of the SBL and Verbal and Full Scales when obtained from an eighteen - nineteen year age group, such as employed in this study; however, one cannot be confident of

comparability between the SBL and Verbal and Full Scale IQ's that are obtained on single individuals within the group. Apparently, this is due to the fact that the WAIS is a less variable measure, while the SBL has greater variability with an older group.

E. Distribution of WAIS and SBL IQ's.

The various intelligence levels proposed by Wechsler and the percentage of individuals within each are presented in Table 9. The last column of this table indicates the percentage expected according to the normal curve.

In order to determine whether the percentage of individuals within each level could probably have arisen from a normally distributed sample, a chi square test was made (18, p. 284). A chi square of 1.260 was obtained on the WAIS distribution indicating that this distribution was not significantly different from a normal distribution at the one per cent level. A chi square of 48.78 was obtained on the SBL distribution indicating that this distribution was significantly different from a normal distribution at the one per cent level.

It was mentioned above that the variability of the SBL was probably due to two reasons. One reason was that duller subjects tend to obtain lower IQ's on the SBL while brighter subjects tend to obtain higher IQ's. The greater percentage of subjects in the 79 and below IQ group and in the 120 and above IQ group would seem to confirm this assumption. It would appear that for this reason the distribution of SBL IQ's did not conform to normal curve expectancy. Since the WAIS conforms to normal curve expectancy while the SBL does not, and since the variability of the WAIS is approximately that of the general population, The WAIS appears to be a better test, which would confirm

Table 9

DISTRIBUTION OF WAIS FULL SCALE, AND SBL IQ'S

(Wechsler's Classification)

Intelligence Level	WAIS N = 100	SBL N = 100	Normal Curve Expectancy
	Percentage	Percentage	Percentage
69 and below	2	4	2.2
70 - 79	6	13	6.7
80 - 89	15	5	16.1
90 -109	51	36	50.0
110 -119	18	16	16.1
120 -129	9	17	6.7
130 and above	1	9	2.2

Wechsler's inference (61, p. 22).

CHAPTER V

SUMMARY AND CONCLUSIONS

The primary purpose of this study was to determine whether the WAIS was a reliable instrument of measurement when administered to a population sample other than the standardization group. A secondary purpose was to compare the results of the WAIS with those of another intelligence scale, the SBL, to determine the relationship of the two scales.

In order to achieve these purposes, the WAIS and SBL were administered to one hundred individuals between the ages of eighteen and twenty years of age. These subjects were selected according to a stratified sampling plan based on an analysis of the 1950 United States Census. After the selection and testing were completed, it was found that except for the higher educational level of female subjects, the testees appeared to be a representative sampling of the national eighteen-nineteen year age group.

The split-half method of estimating reliability was used to determine the reliability coefficients of each subtest and IQ scale of the WAIS. It was found that the reliability coefficients obtained on the sample used in this study did not differ significantly from those reported by Wechsler on the standardization group. On the basis of the present study, it can be assumed that the reliability coefficients of WAIS subtests and IQ scales as published by Wechsler are stable for an eighteen-nineteen year age group.

The correlation of the WAIS IQ scales with the SBL IQ conformed to

expectation. The Verbal and Full Scale IQ correlated highly with the SBL IQ, while the Performance Scale IQ correlated to a lesser extent. Coefficients of correlation between the SBL and Verbal, Performance, and Full Scale IQ's were .92, .81, and .92 respectively. It was noted, however, that the correlation obtained in this study between Performance Scale and SBL IQ's was higher than those reported in previous studies correlating W-B Performance Scale IQ's with those of the SBL. It was also significantly higher than that reported by Wechsler for the comparison study he did between the WAIS Performance Scale and SBL IQ's (.69). This result, however, could be a function of the restricted population used in his study. On the basis of the higher than expected correlation between WAIS Performance Scale and SBL IQ's found in the present study, it was hypothesized that the higher correlation was due to changes in the WAIS Performance scale which had increased its verbal content. The validity of this hypothesis will be determined by future research comparing the WAIS and SBL.

The mean IQ and standard deviation obtained in the present study for each WAIS IQ scale was found not to be significantly different from the general population mean and standard deviation. Although the mean IQ of the SBL was found not significantly different from the general population mean, the SBL standard deviation was significantly different at the one per cent level. It was hypothesized that the greater variability of the SBL was because duller subjects tend to obtain lower IQ's on the SBL while brighter subjects tend to obtain higher IQ's on the SBL. The greater percentage of subjects in the lower and higher IQ groups would tend to increase SBL variability. Another possible reason for the greater variability of the SBL was that this test was

standardized on children and could not be expected to have a variability comparable with the WAIS which was standardized on a sample similar to that used in the present study.

The chi square test was used to determine whether the distribution of IQ's for both tests could have arisen from a normally distributed sample. It was found that the distribution of WAIS IQ's did not differ significantly. The SBL IQ's, however, did have a significantly different distribution from a normally distributed sample at the one per cent level of significance. This difference appeared to be due to the greater variability of the SBL IQ's resulting from the reasons mentioned above.

It can be concluded from this study that the reliability coefficients reported by Wechsler for the subtests and IQ scales of the WAIS are valid for an eighteen and nineteen year old population. A second conclusion is that the WAIS, in contrast to the SBL, has less variability and the results of individual administrations to eighteen and nineteen year old subjects can be accepted with confidence in relation to the published reliability coefficients of each subtest and IQ scale. The SBL has a significantly different variability when administered to an eighteen-nineteen year old population and the results of individual tests may be highly unreliable.

BIBLIOGRAPHY

1. Alderice, E.T., & Butler, A.J. An analysis of the performance of mental defectives on the Revised Stanford-Binet and the Wechsler-Bellevue Intelligence Scale. Amer. J. ment. Def., 1952, 56, 609-614.

2. Anastasi, Anne. Psychological Testing. N.Y.: Macmillan Co., 1954.

3. Anderson, E.E., et. al. A comparison of the Wechsler-Bellevue, Revised Stanford-Binet, and American Council of Education Tests at the college level. J. Psychol., 1942, 14, 317-326.

4. Benton, A.L., Weider, A., & Blauvelt, J. Performance of Adult Patients on the Bellevue Intelligence Scales and the Revised Stanford-Binet. Psychia. Quart., 1941, 15, 802-806.

5. Botwinick, J. Wechsler-Bellevue split-half subtest reliabilities: differences in age and mental status. J. consult. Psychol., 1953, 17, 225-228.

6. Burik, T.E. Investigation of the learning involved in the digit symbol subtest of the Wechsler-Bellevue Intelligence Scale. J. Psychol., 1950, 30, 32-42.

7. Cole, D., & Webela, Lois. Comparison Data on the Wechsler-Bellevue and the Wechsler Adult Intelligence Scale. J. clin. Psychol., 1956, 12, 198-199.

8. Cronbach, L.J. Essentials of Psychological Testing. N.Y.: Harper & Brothers, 1949.

9. Dana, R.H. A comparison of four verbal subtests on the Wechsler-Bellevue, Form I, and the WAIS. J. clin. Psychol., 1957, 13, 70-71.

10. Derner, G.F., Aborn, M., & Canter, A.H. The reliabilities of the Wechsler-Bellevue subtests and scales. J. consult. Psychol., 1950, 14, 172-179.

11. Doppelt, J.E. Estimating the full scale score on the Wechsler Adult Intelligence Scale from scores on four subtests. J. consult. Psychol., 1956, 20, 63-66.

12. Doppelt, J.E., & Wallace W. Standardization of the Wechsler Adult Intelligence Scale for older persons. J. abnorm. soc. Psychol., 1955, 51, 312-330.

13. Gilhooly, F.M. Wechsler-Bellevue reliability and validity of certain diagnostic signs of the neuroses. J. consult. Psychol., 1950, 14, 81-87.

14. Goldfarb, W. Adolescent performance in the Wechsler-Bellevue Intelligence Scales and the Revised Stanford-Binet examination, Form L. J. educ. Psychol., 1944, 35, 503-507.

15. Goolishian, H.A., & Ramsay, Rose. The Wechsler-Bellevue, Form I, and the Wechsler Adult Intelligence Scale: A Comparison. J. clin. Psychol., 1956, 12, 147-150.

16. Guertin, W.H. Mental growth in pseudo-feeblemindedness. J. clin. Psychol., 1949, 5, 404-418.

17. Guertin, W.H., Frank, G.H., & Rabin, A.I. Research with the Wechsler-Bellevue Intelligence Scale. Psychol. Bull., 1956, 53, 235-257.

18. Guilford, J.P. Fundamental Statistics in Psychology and Education. N.Y.: McGraw-Hill, 1950.

19. Gulliksen, H. Theory of Mental Tests. N.Y. Wiley, 1950.

20. Gurvitz, M.S. Some defects of the Wechsler-Bellevue. J. consult. Psychol., 1952, 16, 124-126.

21. Hall, Julia C. Correlation of a modified form of Raven's Progressive Matrices (1938) with the Wechsler Adult Intelligence Scale. J. consult. Psychol., 1957, 20, 23-26.

22. Halpern, Florence. A comparison of the Revised Stanford L and the Bellevue Adult Intelligence Test as clinical instruments. Psychiat. Quart. Suppl., 1942, 16, 206-211.

23. Hamister, R.C. Test-retest reliability of the Wechsler-Bellevue. J. consult. Psychol., 1949, 13, 39-44.

24. Helmick, J.S. Reliability or variability? J. consult. Psychol., 1952, 16, 154-155.

25. Himelstein, P. Evaluation of an abbreviated WAIS in a psychiatric hospital. J. clin. Psychol., 1957, 13, 68-69.

26. Jastak, J. An item analysis of the Wechsler-Bellevue tests. J. consult. Psychol., 1950, 14, 88-94.

27. Jones, H.G. The evaluation of the significance of differences between scale scores on the WAIS: The perpetuation of a fallacy. J. consult. Psychol., 1956, 20, 319-320.

28. Kelley, T.L. Fundamentals of Statistics. Cambridge: Harvard Univ. Press, 1947.

29. Kitzinger, Helen, & Blumberg, E. Supplementary guide for administering and scoring the Wechsler-Bellevue Intelligence Scale, Form I. Psychol. Monogr., 1951, 65, No. 2 (Whole No. 319).

30. 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.

31. Loutitt, C.M., & Browne, C.G. The use of psychometric instruments in psychological clinics. J. consult. Psychol., 1947, 11, 49-54.

32. Luchins, A.S. Teaching experimental psychology to clinical students. Amer. Psychologist, 1949, 4, 48-51.

33. Luchins, A.S., & Luchins, Edith. Effects of varying the administration of the digit symbol subtest of the Wechsler-Bellevue Intelligence Scale. J. gen. Psychol., 1953, 49, 125-142.

34. McNemar, Q. The Wechsler Adult Intelligence Scale. A Review. J. consult Psychol., 1956, 20. 159.

35. Mech, E. Item analysis and discriminative value of selected Wechsler-Bellevue subtests. J. educ. Res., 1953, 47, 241-260.

36. Mitchell, M.B. Performance of mental hospital patients in the Wechsler-Bellevue and the Revised Stanford-Binet, Form L. J. educ. Psychol., 1942, 33, 538-544.

37. Norman, R.D. Sex differences and other aspects of young superior adult performance on the Wechsler-Bellevue. J. consult. Psychol., 1953, 17, 411-418.

38. Olin, T.D. & Resnikoff, M. The use of Doppelt's short form of the Wechsler Adult Intelligence Scale with psychiatric patients. J. consult. Psychol., 1957, 21, 27-28.

39. Patterson, C.H. A Guide for Counselors. Springfield, Ill.: Charles C. Thomas, 1953.

40. Rabin, A.I. Fluctuations in the mental level of schizophrenic patients. Psychiat. Quart., 1944, 18, 78-91.

41. Rabin, A.I. Test constancy and variation in the mentally ill. J. gen. Psychol., 1944, 31, 231-239.

42. Rabin, A.I. The use of the Wechsler-Bellevue scales with normal and abnormal persons. Psych. Bull., 1945, 42, 410-422.

43. Rabin, A.I., & Guertin, W.H. Research with the Wechsler-Bellevue Test: 1945-1950. Psych. Bull., 1951, 48, 211-248.

44. Russell, G.E. Wechsler-Bellevue Vocabulary subtest items: revised order of words. Med. Tech. Bull., 1954, 5, 143-148.

45. Sartain, A.Q. A comparison of the New Revised Stanford-Binet, the Bellevue Scale, and certain group tests of intelligence. J. soc. Psychol., 1946, 23, 237-239.

46. Shafer, R. The Wechsler Adult Intelligence Scale: A Review. J. consult. Psychol., 1956, 20, 157-159.

47. Stacey, C.L., & Levin, Janice. Performance of retarded individuals on the Stanford-Binet and Wechsler-Bellevue Intelligence Scales. Amer. J. ment. Defic., 1950, 55, 123-152.

48. Watson, R.I. The Clinical Method in Psychology. N.Y.: Harper and Brothers, 1951.

49. Webb, W.B., & de Haan, H. Wechsler-Bellevue split-half reliabilities in normals and schizophrenics. J. consult. Psychol., 1951, 15, 68-71.

50. Wechsler, D. The Measurement of Adult Intelligence (3rd ed.) Baltimore: Williams and Wilkins, 1944.

51. Wechsler, D. Manual for the Wechsler Adult Intelligence Scale. N.Y.: Psychological Corporation, 1955.

52. Weider, A., Levi, J., & Risch, F. Performance of problem children on the Wechsler-Bellevue Intelligence Scales and the Revised Stanford-Binet. Psychiat. Quart., 1943, 17, 695-701.

APPENDIX I

Table 10

STANDARD DEVIATIONS OF ODD (g) AND EVEN HALVES (G)

OF ELEVEN WAIS SUBTESTS

Test	Odd (g)	Even (G)
1. Information	2.54	3.05
2. Comprehension	2.66	2.72
3. Arithmetic	1.82	1.87
4. Similarities	2.61	2.57
5. Digit Span	2.48	2.47
6. Vocabulary	8.42	8.53
7. Digit Symbol	6.23	6.32
8. Picture Completion	2.04	2.38
9. Block Design	4.96	5.82
10. Picture Arrangement	3.78	3.56
11. Object Assembly	5.63	6.35

APPENDIX II

Table 11

PRODUCTS OF STANDARD DEVIATIONS OF ODD HALVES (g) WITH STANDARD DEVIATION OF EVEN HALVES (G) OF ELEVEN WAIS SUBTESTS ($S_g S_G$)

		EVEN HALVES										
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI
ODD HALVES	1	7.70	6.91	4.75	6.53	6.27	21.67	16.05	6.05	14.78	9.04	16.08
	2	8.06	7.24	4.97	6.84	6.57	22.69	16.81	6.33	15.48	9.47	16.84
	3	5.51	4.95	3.40	4.68	4.50	15.52	11.50	4.33	10.59	6.48	11.52
	4	7.91	7.10	4.88	6.71	6.45	22.26	16.50	6.21	15.19	9.29	16.52
	5	7.61	6.75	4.64	6.37	6.13	21.15	15.67	5.90	14.43	8.83	15.70
	6	25.57	22.90	15.75	21.64	20.80	71.99	53.34	20.09	49.12	30.05	53.43
	7	18.88	16.95	11.65	16.01	15.39	53.14	39.37	14.83	36.26	22.18	39.44
	8	6.18	5.56	3.81	5.24	5.04	17.40	12.89	4.86	11.87	7.26	12.91
	9	15.03	13.49	9.28	12.75	12.25	42.31	31.35	11.80	28.87	17.66	31.40
	10	11.45	10.28	7.07	9.71	9.34	32.24	23.89	9.00	22.00	13.46	23.93
	11	11.00	9.87	6.79	9.33	8.97	30.96	22.94	8.64	21.13	12.92	22.98

APPENDIX III

Table 12

COEFFICIENTS OF CORRELATION BETWEEN ODD HALVES (g) AND EVEN HALVES (G) OF ELEVEN WAIS SUBTESTS (r_{gG})

ODD HALVES	EVEN HALVES										
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
1	.793	.660	.622	.527	.392	.699	.433	.581	.470	.508	.412
2	.569	.651	.564	.542	.300	.635	.453	.604	.607	.539	.541
3	.591	.613	.689	.499	.451	.608	.440	.522	.466	.506	.396
4	.571	.638	.513	.702	.388	.671	.450	.489	.487	.512	.364
5	.382	.370	.385	.355	.546	.384	.255	.513	.340	.350	.235
6	.755	.666	.692	.651	.396	.909	.400	.588	.513	.572	.459
7	.407	.482	.406	.460	.395	.449	.908	.412	.504	.351	.384
8	.579	.598	.501	.445	.337	.572	.500	.612	.553	.574	.556
9	.484	.592	.548	.458	.436	.505	.515	.528	.773	.511	.648
10	.367	.337	.340	.390	.346	.393	.356	.309	.277	.487	.299
11	.376	.401	.441	.432	.318	.444	.469	.461	.571	.509	.555

APPENDIX IV

Table 13

PRODUCTS OF STANDARD DEVIATIONS OF ODD HALVES AND EVEN HALVES (TABLE 11) WITH COEFFICIENTS OF CORRELATION OF ODD HALVES AND EVEN HALVES (TABLE 12) ($r_{gG} S_g S_G$)

		EVEN HALVES										
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI
ODD HALVES	1	6.11	4.56	2.95	3.44	2.47	15.15	6.95	3.52	6.95	4.59	6.62
	2	4.59	4.71	2.80	3.71	1.97	14.41	7.61	3.82	9.40	5.10	9.11
	3	3.26	3.03	2.34	2.34	2.03	9.44	5.06	2.26	4.93	3.28	4.56
	4	4.52	4.53	2.50	4.71	2.50	14.94	7.42	3.04	7.40	4.76	6.01
	5	2.87	2.50	1.79	2.66	3.35	8.12	4.00	1.85	4.91	3.09	3.69
	6	19.31	15.25	10.90	14.09	8.24	65.44	21.34	11.81	25.20	17.19	24.52
	7	7.68	8.17	4.73	7.36	6.06	23.86	35.75	6.11	18.28	7.79	15.14
	8	3.58	3.32	2.25	2.33	1.70	9.95	6.44	2.97	6.56	4.17	7.18
	9	7.27	7.99	5.08	5.84	5.34	21.37	16.15	6.23	22.32	9.02	20.35
	10	4.20	3.46	2.40	3.79	3.23	12.67	8.50	2.78	6.09	6.56	7.16
	11	4.14	3.96	2.99	4.03	2.85	13.75	10.76	3.98	12.07	6.58	12.75

APPENDIX V

Table 14

PRODUCTS OF STANDARD DEVIATIONS OF EVEN HALVES (G) WITH STANDARD DEVIATIONS OF EVEN HALVES (G) OF WAIS SUBTESTS ($S_G S_G$)

		EVEN HALVES										
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI
EVEN HALVES	I	9.18	8.24	5.67	7.79	7.48	25.85	19.15	7.21	17.63	10.79	19.18
	II	8.24	7.40	5.09	6.99	6.72	23.20	17.19	6.47	15.83	9.68	17.22
	III	5.67	5.09	3.50	4.81	4.62	15.95	11.82	4.45	10.88	6.66	11.84
	IV	7.79	6.99	4.81	6.60	6.35	21.92	16.24	6.12	14.96	9.15	16.27
	V	7.48	6.72	4.62	6.35	6.10	21.07	15.61	5.89	14.38	8.79	15.64
	VI	25.85	23.20	15.95	21.92	21.07	72.76	53.91	20.30	49.64	30.37	53.99
	VII	19.15	17.19	11.82	16.24	15.61	53.91	39.94	15.04	36.78	22.50	40.01
	VIII	7.21	6.47	4.45	6.12	5.89	20.30	15.04	5.66	13.85	8.47	15.07
	IX	17.63	15.83	10.88	14.96	14.38	49.64	36.78	13.85	33.87	20.72	36.84
	X	10.79	9.68	6.66	9.15	8.79	30.37	22.50	8.47	20.72	12.67	22.53
	XI	19.18	17.22	11.84	16.27	15.64	53.99	40.01	15.07	36.84	22.53	40.07

APPENDIX VI

Table 15

COEFFICIENTS OF CORRELATION BETWEEN EVEN HALVES OF WAIS SUBTESTS (r_{GG})

		EVEN HALVES										
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI
EVEN HALVES	I	1.00	.600	.696	.539	.376	.757	.349	.627	.448	.505	.331
	II	.600	1.00	.634	.579	.391	.724	.436	.598	.558	.510	.503
	III	.696	.634	1.00	.450	.396	.686	.363	.635	.491	.409	.379
	IV	.539	.579	.450	1.00	.308	.662	.385	.418	.475	.455	.400
	V	.376	.391	.396	.308	1.00	.354	.378	.361	.335	.257	.282
	VI	.757	.724	.686	.662	.354	1.00	.403	.581	.542	.542	.527
	VII	.349	.436	.363	.385	.378	.403	1.00	.388	.476	.335	.412
	VIII	.627	.598	.635	.418	.361	.581	.388	1.00	.554	.533	.565
	IX	.448	.558	.491	.475	.335	.542	.476	.554	1.00	.502	.679
	X	.505	.510	.409	.455	.257	.542	.335	.533	.502	1.00	.541
	XI	.331	.503	.379	.400	.282	.527	.412	.565	.679	.541	1.00

APPENDIX VII

Table 16

PRODUCTS OF STANDARD DEVIATIONS OF EVEN HALVES AND STANDARD DEVIATIONS OF EVEN HALVES (TABLE 14) WITH COEFFICIENTS OF CORRELATION BETWEEN EVEN HALVES OF WAIS SUBTESTS (TABLE 15) ($S_G S_G r_{GG}$)

	EVEN HALVES										
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
I	9.18	4.94	3.95	4.20	2.81	19.57	6.68	4.52	7.90	5.45	6.35
II	4.94	12.39	3.23	4.05	2.63	16.90	7.49	3.87	8.83	4.94	8.66
III	3.95	3.23	4.88	2.16	1.83	10.94	4.29	2.83	5.34	2.72	4.49
IV	4.20	4.05	2.16	7.95	1.96	14.51	6.25	2.56	7.11	4.16	6.51
V	2.81	2.63	1.83	1.96	9.73	7.46	5.90	2.13	4.82	2.26	4.41
VI	19.57	16.80	10.94	14.51	7.46	72.76	21.73	11.79	26.90	16.46	28.45
VII	6.68	7.49	4.29	6.25	5.90	21.73	39.94	5.84	17.51	7.54	16.48
VIII	4.52	3.87	2.83	2.56	2.13	11.79	5.84	5.66	7.67	4.51	8.51
IX	7.90	8.83	5.34	7.11	4.82	26.90	17.51	7.67	33.87	10.40	25.01
X	5.45	4.94	2.72	4.16	2.26	16.46	7.54	4.51	10.40	12.67	12.19
XI	6.35	8.66	4.49	6.51	4.41	28.45	16.48	8.51	25.01	12.19	40.07

EVEN HALVES

APPENDIX VIII

Table 17

PRODUCTS OF STANDARD DEVIATIONS OF ODD HALVES (g) WITH STANDARD DEVIATIONS OF ODD HALVES (g) OF WAIS SUBTESTS ($S_g S_g$)

ODD HALVES	1	2	3	4	5	6	7	8	9	10	11
1	6.45	6.76	4.63	6.63	6.30	21.44	15.82	5.18	12.60	9.60	9.22
2	6.76	7.08	4.84	6.94	6.60	22.40	16.57	5.43	13.19	10.05	9.66
3	4.62	4.84	3.31	4.75	4.51	15.32	11.34	3.71	9.03	6.88	6.61
4	6.63	6.94	4.75	6.81	6.47	21.98	16.26	5.32	12.95	9.87	9.47
5	6.30	6.60	4.51	6.47	6.15	20.88	15.45	5.06	12.30	9.37	9.00
6	21.44	22.40	15.32	21.98	20.88	71.23	52.58	17.22	41.86	31.90	30.64
7	15.82	16.57	11.34	16.26	15.45	52.58	38.81	12.71	30.90	23.55	22.61
8	5.18	5.43	3.71	5.32	5.06	17.22	12.71	4.16	10.12	7.71	7.41
9	12.60	13.19	9.03	12.95	12.30	41.86	30.90	10.12	24.60	18.75	18.00
10	9.60	10.05	6.88	9.87	9.37	31.90	23.55	7.71	18.75	14.29	13.72
11	9.22	9.66	6.61	9.47	9.00	30.64	22.61	7.41	18.00	13.72	13.18

APPENDIX IX

Table 18

COEFFICIENTS OF CORRELATION BETWEEN ODD HALVES OF WAIS SUBTESTS (r_{gg})

		ODD HALVES										
		1	2	3	4	5	6	7	8	9	10	11
ODD HALVES	1	1.00	.651	.624	.591	.311	.743	.466	.537	.498	.389	.390
	2	.651	1.00	.555	.598	.380	.676	.495	.642	.582	.457	.558
	3	.624	.555	1.00	.533	.338	.623	.432	.585	.566	.425	.438
	4	.591	.598	.533	1.00	.360	.649	.516	.468	.480	.358	.508
	5	.311	.380	.338	.360	1.00	.392	.324	.331	.391	.286	.272
	6	.743	.676	.623	.649	.392	1.00	.438	.602	.509	.468	.470
	7	.466	.495	.432	.516	.324	.438	1.00	.488	.530	.356	.475
	8	.537	.642	.585	.468	.331	.602	.488	1.00	.605	.381	.528
	9	.498	.582	.566	.480	.391	.509	.530	.605	1.00	.362	.593
	10	.389	.457	.425	.358	.286	.468	.356	.381	.362	1.00	.315
	11	.390	.558	.438	.508	.272	.470	.475	.528	.593	.315	1.00

APPENDIX X

Table 19

PRODUCTS OF STANDARD DEVIATIONS OF ODD HALVES AND STANDARD DEVIATIONS OF ODD HALVES (TABLE 17) WITH COEFFICIENTS OF CORRELATION BETWEEN ODD HALVES OF WAIS SUBTESTS (TABLE 18) ($S_g S_g r_{gg}$)

	ODD HALVES										
	1	2	3	4	5	6	7	8	9	10	11
1	6.45	4.40	2.88	3.92	1.96	15.93	7.37	2.78	6.27	3.73	3.60
2	4.40	7.08	2.69	4.15	2.51	15.14	8.20	3.49	7.68	4.59	5.39
3	2.88	2.69	3.31	2.53	1.52	9.54	4.90	2.17	5.11	2.92	2.90
4	3.92	4.15	2.53	6.81	2.33	14.27	8.39	2.49	6.22	3.53	4.81
5	1.96	2.51	1.52	2.33	6.15	8.18	5.01	1.67	4.81	2.68	2.45
6	15.93	15.14	9.54	14.27	8.18	71.23	23.03	10.37	21.31	14.93	14.40
7	7.37	8.20	4.90	8.39	5.01	23.03	38.81	6.20	16.38	8.38	10.74
8	2.78	3.49	2.17	2.49	1.67	10.37	6.20	4.16	6.12	2.94	3.91
9	6.27	7.68	5.11	6.22	4.81	21.31	16.38	6.12	24.60	6.79	10.67
10	3.73	4.59	2.92	3.53	2.68	14.93	8.38	2.94	6.79	14.29	4.32
11	3.60	5.39	2.90	4.81	2.45	14.40	10.74	3.91	10.67	4.32	13.18

ODD HALVES

APPROVAL SHEET

The dissertation submitted by Fred D. Whelan, has been read and approved by five members of the Department of Psychology.

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated, and that the dissertation is now given final approval with reference to content, form, and mechanical accuracy.

The dissertation is therefore accepted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

November 18, 1957
Date

Frank J. Noble
Signature of Adviser