A COMPARISON OF INTELLIGENCE QUOTIENTS
FOR READERS AND NON-READERS USING
THE LEITER INTERNATIONAL PERFORMANCE SCALE AND
THE 1937 REVISION OF THE STANFORD-BINET

by

Barbara Ann Wilson
B. A. (Sacramento State College) 1957

THESIS

Submitted in partial satisfaction of
the requirements for the degree of

MASTER OF SCIENCE IN APPLIED PSYCHOLOGY
AT THE
SACRAMENTO STATE COLLEGE

Approved:

Date 12/7/57
ACKNOWLEDGMENTS

The writer is grateful to the following people for their help in the preparation of this thesis:

The Committee: Dr. R. A. Good, Chairman, Dr. Harry Aron and Dr. J. D. Lucas; Mrs. Mildred Lonsdale, Psychologist, North Sacramento Unified School District; the principals and teachers of the North Sacramento School District who were kind enough to set up appointments and release the children for testing.

To my husband, Robert Wilson, and my sister, Etta Marie Cullivan, who provided aid and comfort when they were sorely needed, sincerest appreciation.
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. THE PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td>II. REVIEW OF ORIGIN AND RATIONALES BEHIND THE LEITER INTERNATIONAL PERFORMANCE SCALE AND THE STANFORD-BINET TEST, 1937 REVISION</td>
<td>3</td>
</tr>
<tr>
<td>Leiter International Performance Scale</td>
<td>3</td>
</tr>
<tr>
<td>Stanford-Binet Test of General Intelligence</td>
<td>7</td>
</tr>
<tr>
<td>The original form of the test</td>
<td>7</td>
</tr>
<tr>
<td>The Terman-Merrill revision</td>
<td>10</td>
</tr>
</tbody>
</table>

Studies of the Leiter and the Binet [in]

<table>
<thead>
<tr>
<th>Year</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1952</td>
<td>L. Beverly and G. Bensberg</td>
<td>12</td>
</tr>
<tr>
<td>1951</td>
<td>Robert T. Glenn</td>
<td>13</td>
</tr>
<tr>
<td>1950</td>
<td>Trent E. Bessent</td>
<td>16</td>
</tr>
<tr>
<td>1950</td>
<td>Guy L. Bond and Leo C. Far</td>
<td>16</td>
</tr>
<tr>
<td>1949</td>
<td>Jack Mathews</td>
<td>17</td>
</tr>
<tr>
<td>1946</td>
<td>Natalie Darcy</td>
<td>18</td>
</tr>
<tr>
<td>1946</td>
<td>Charlotte Madeley</td>
<td>19</td>
</tr>
<tr>
<td>1941</td>
<td>Ruth Williams</td>
<td>20</td>
</tr>
<tr>
<td>1939</td>
<td>Morris Krugman</td>
<td>22</td>
</tr>
</tbody>
</table>

Summary                                            | 23   |
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>III. THE ASSUMPTIONS, HYPOTHESIS, AND PROCEDURE</td>
<td>24</td>
</tr>
<tr>
<td>The Assumptions</td>
<td>24</td>
</tr>
<tr>
<td>The Hypothesis</td>
<td>24</td>
</tr>
<tr>
<td>The Procedure</td>
<td>26</td>
</tr>
<tr>
<td>IV. CONCLUSIONS, DISCUSSION, AND SUMMARY</td>
<td>31</td>
</tr>
<tr>
<td>Conclusions</td>
<td>31</td>
</tr>
<tr>
<td>Discussion</td>
<td>31</td>
</tr>
<tr>
<td>Summary</td>
<td>32</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>35</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>38</td>
</tr>
<tr>
<td>TABLE</td>
<td>PAGE</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>I. Comparison of Average Chronological Age and Average Mental Age</td>
<td>6</td>
</tr>
<tr>
<td>II. Evidences for the Reliability and Validity of the Leiter International Performance Scale</td>
<td>8</td>
</tr>
<tr>
<td>III. The Binet-Leiter Correlation</td>
<td>15</td>
</tr>
<tr>
<td>IV. The Intercorrelations of the Relationship of the Leiter to the Stanford-Binet</td>
<td>21</td>
</tr>
<tr>
<td>V. Reader and Non-reader IQ Scores Based on the Leiter and Binet</td>
<td>29</td>
</tr>
<tr>
<td>FIGURE</td>
<td>PAGE</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>1. Example of Leiter Notched Frame</td>
<td>39</td>
</tr>
</tbody>
</table>
CHAPTER I

THE PROBLEM

The problem to be investigated in this study is the constancy of IQ scores of readers and non-readers using the 1937 Revision of the Stanford-Binet Test of General Intelligence,¹ and the Leiter International Performance Scale, 1948 Revision.²

Inasmuch as the Binet is a scale which is weighted heavily toward verbal content (6, p. 114), it is possible that it would result in a significantly different estimate of intelligence for verbally handicapped children than would the Leiter, which appears to be a purely non-verbal test.

This problem becomes important when children are tested to obtain an estimate of capacity to learn. If the child is verbally handicapped and if the Binet tends to penalize this handicap, an incorrect estimate of the child's capacity is obtained. A test, then, which would not depend on verbal ability would perhaps give a better estimate of "general intelligence."³ The Leiter scale was originally designed to be

¹ Also referred to as the Binet.
² Also referred to as the Leiter.
³ In this study "general intelligence" as measured by the Binet is defined as "scholastic aptitude, readiness to do the sort of tasks required in school" (6, p. 114).
just such a test. Leiter (10) hoped to measure the same intel-
lectual ability as was measured by the Binet, without us-
ing verbal tasks.

The purpose of this study is to attempt to see what
type of differences would result from the comparison of the
Leiter and the Binet scores, using reading and non-reading
children.
CHAPTER II

REVIEW OF ORIGIN AND RATIONALE BEHIND
THE LEITER INTERNATIONAL PERFORMANCE SCALE
AND THE STANFORD-BINET TEST, 1937 REVISION

This chapter will attempt to review the origin and rationale behind the Leiter and the Binet. The pertinent studies comparing these scales will be included.

I. LEITER INTERNATIONAL PERFORMANCE SCALE

The Leiter was first published in 1940. The 1940 form of the test was based upon a 1938 version which Dr. Leiter developed in Hawaii, using Japanese and Hawaiian children as his subjects. He did not use any Caucasian children because he felt that the Caucasian population was too highly selected at that time. At the time, the Caucasian population was almost entirely composed of professional people and their families. Thus, the population was not representative of the general white population of the United States.

Dr. Leiter's work in Hawaii, and his previous experiments in non-verbal test construction (his short 1936 scale which he claimed correlated .79 with the 1916 Binet), led him to assume that,
if the language factor in mental tests could be eliminated, the differences in native intelligence between children of various races, which had been found with previous investigators using tests which required the use of language, would no longer be found to exist (10, p. 10).

(The 1938 scale was composed of 56 tests in groups of four per year from year 3 to year 10, and even years from 12 to 22. Each test was worth three months mental age credit up to year 10, six months from years 12 to 16, nine months at year 18 and twelve months at years 20 and 22. This scale was the one administered to 764 Japanese and Chinese children.)

In the fall of 1938 Dr. Leiter returned to the mainland and administered his test to 280 middle-class white children. The difference between the performance of the two groups was great enough, he felt, to demonstrate that the test was unsuitable for use with a Caucasian group. From Leiter's study (10, p. 13) it is clear that he found that the chronological ages and mental ages of his Caucasian group were so disparate as to invalidate the test for this group. Consequently, he abandoned his plans for developing an international test of performance and he undertook to develop a performance scale suitable for use with Caucasians. The revised 1936 scale became the basis for the 1938 scale. The 1938 scale, using 107 cases of American white children, had a correlation coefficient of .51 with the 1916 Binet (15, p. 18).
The major changes in the 1940 scale were (1) the relocation of tests already in the scale, and (2) the addition of new tests suitable for use about the ten year level. The 1940 scale was applied to 280 middle class American white children in California who "were equally distributed between ages 5-0 and 12-11" (10, p. 30). There were thirty-five cases at each year. The reliability of the test was determined by finding the coefficient of reliability by the split-half method and comparing the standard deviation of scores at each age with the S.D. of the revised Stanford-Binet at the same age levels. The coefficient of reliability equalled .89 (the Spearman-Brown correction raised this to .94). The standard deviations are "well below those reported by Terman and Merrill at the same age levels" (10, p. 32).

Further use of the 1940 revision indicated that the test would be easier to administer if tests appeared only at the even year levels beyond year ten.

Dr. Grace Arthur (1) had also prepared a revision of the Leiter up to year twelve, and to simplify administration made some changes in the form of the test. Test XI-2 was substituted for IX-2, XI-1 substituted for X-5, a single frame (see Appendix) was used instead of six of varying lengths, a lighter carrying case was developed and no tests were given at the eleven year level. Children having a mental age up
<table>
<thead>
<tr>
<th>Age level (in years)</th>
<th>Average C.A. (in months)</th>
<th>Average M.A.</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>66.09</td>
<td>72.34</td>
</tr>
<tr>
<td>6</td>
<td>77.05</td>
<td>95.05</td>
</tr>
<tr>
<td>7</td>
<td>89.20</td>
<td>109.62</td>
</tr>
<tr>
<td>8</td>
<td>101.60</td>
<td>127.11</td>
</tr>
<tr>
<td>9</td>
<td>113.42</td>
<td>142.65</td>
</tr>
<tr>
<td>10</td>
<td>126.82</td>
<td>154.34</td>
</tr>
<tr>
<td>11</td>
<td>138.65</td>
<td>164.54</td>
</tr>
<tr>
<td>12</td>
<td>149.68</td>
<td>169.28</td>
</tr>
</tbody>
</table>

*Comparison at each age level from 5 to 12 on 1938 scale when that scale was applied to 280 middle class American white children. Table quoted from Table XI, page 13 of Leiter (10).
to 7.99 can be tested on either the Arthur adaptation or the 1948 Revision. Above that age it is necessary to use the 1948 version of the Leiter.

Dr. Arthur's changes made the Leiter scale for the younger age group a point scale, rather than a mental age scale.

Comparisons of the 1940 and 1948 revision led the author to state that the two tests were measuring the same type of "general intelligence" ($r = .92$). Studies (10, p. 58) have indicated that the test is scale about six months (or 5 points) too low, as compared to the 1937 revision of the Binet. Leiter feels that adding five points to the score will make it comparable with the results of other tests (10, p. 58).

Excerpts from Tables LXI and LX in Leiter's manual (10, p. 60) are included to show the range of correlations obtained for the 1948 Leiter, the Arthur adaptation, and the revised Binet. In the studies Dr. Leiter gives the figures which indicate that his test has a significant correlation with the Binet. These figures are given in Table II.

II. THE STANFORD-BINET TEST OF GENERAL INTELLIGENCE

The original form of the test. Binet designed this test simply as a method by which the intelligence of a child could be estimated. "The method consists in asking the child
TABLE II
EVIDENCES FOR THE RELIABILITY AND VALIDITY OF
THE LEITER INTERNATIONAL PERFORMANCE SCALE*

<table>
<thead>
<tr>
<th></th>
<th>M.A.</th>
<th>I.Q.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leiter</td>
<td>.77</td>
<td>.81</td>
</tr>
<tr>
<td>Arthur adaptation</td>
<td>.81</td>
<td>.79</td>
</tr>
</tbody>
</table>


**The number of cases used to determine these figures was not given.
some precise questions and asking him to perform some simple
experiments, these questions and experiments are called
tests" (4, p. 7).

The original French form of the test was not a great
deal different from the Terman and Merrill revisions. Some
of the items are the same except that the reference point is
the United States. / For example, in the original test, year
15-5, there was this item:

1. A woman walking in the forest of Fontainbleau
stopped suddenly, dreadfully frightened, hurried
to the nearest policeman and told him that she had
just seen hanging to the limb of a tree a ----a
what? (4, p. 50)

Almost the identical wording appears in year XIII-4 of the
1937 Revision, Form L.

The original form of the scale appeared to be a valid
and reliable measure over the intermediate age range, but it
was defective at both extremes (13, p. 3). From age 5 to 10
the scale was accurate, but it became progressively less use-
ful with older subjects. Many items had low validity, were
difficult to score, and susceptible to coaching. The instruc-
tions lacked precision and there was no alternate form for
comparison of the scores and retesting. The revision pro-
vided two equivalent scales. Form L is more similar to the
old scale but both forms were designed to be equal. The new
scales cover a wider range, are more accurately standardized,
give a larger sample of abilities, and the procedures are better defined (13, p. 3).

The first scale had 54 tests, the first revision 90, and the 1937 revision had 129 in each form. At the lower end more non-verbal items were used, and the upper end items are mostly of a conceptual nature.

The Terman-Merrill revision. Terman and Merrill attempted to obtain a representative sample of white children from the ages of two to eighteen for their standardization group. The sample was composed of one hundred children at each half-year level below 6, two hundred at each age between 6 and 14, and one hundred from 15 to 18. The same sample was used for both forms of the scale to insure equivalence of the forms.

The test items in the revised form were selected from reviews of other intelligence tests. The items were experimentally checked by Stanford graduate students. The most promising material was used as a preliminary test on one thousand students living in the vicinity of Stanford University.

After this tryout, the items were accepted for use if they met the following requirements: (1) validity (degree not reported), (2) ease and objectivity of scoring, (3) practical considerations such as time economy, interest to the subject, and need for variety (13, p. 9). Validity was determined by (1) increases in the percentages passing from one
mean age (or mental age) to the next, and (2) a weight based on the ratio of the difference to the standard error of the difference between the mean age (or mental age) of the subjects passing the test and the subjects failing it (13, p. 9).

Two provisional scales were used, Form L containing 209 tests and Form M containing 199. Each subject was administered both forms between one day and one week apart. Half of the sample received Form L first and half Form M first. Total time for administration was approximately three hours.

1. Geographical location of the subjects. A total of 3,184 subjects were selected from seventeen different communities in eleven states. These states were California, Nevada, New York, Colorado, Kansas, Virginia, Vermont, Texas, Minnesota, Indiana and Kentucky.

2. Social group. The socio-economic status was estimated by the Sims questionnaire and other rating media. An attempt was made to have approximately the same percentage of people in the sample as were found in each grouping.

3. Schools. Schools of average social status were selected. In each school all the children between 6 and 14 were tested who were within one month of a birthday. The sample of older children was the most difficult to locate.
A sample of cases who were not in school was included. As with the pre-schoolers, most of these were siblings of the children tested who were in school.

4. Nationality. All the subjects were American born, white children. In 50 per cent of the cases both parents were born in the United States.

After all the tests were administered they were checked and rescored. Lists of satisfactory and unsatisfactory responses were made up for each item. The important task at this point was to eliminate the less satisfactory tests and arrange the tests so the mean chronological age would be as close as possible to the mean mental age, so the IQ would be as close as possible to 100.

III. STUDIES OF THE LEITER AND THE BINET
(Chronological order)

1952. L. Beverly and G. Bensberg (3). This study on a group of fifty mental defectives gave a Pearson r of .621 between the Leiter and the Binet. The subjects ranged in age from 6 years, 11 months to 16 years, 2 months. They were tested on the Binet, Form L, the Leiter, and the Cornell-Coxe. The Leiter and the Cornell-Coxe correlated .818 which
was expected because both tests are performance tests, and "measure the same factors." The Cornell-Coxe correlated .670 with the Binet. An analysis of variance gave significant differences between the three tests, the Leiter scoring below both the Binet and the Cornell-Coxe. (The Leiter did not appear to be a more valid instrument for this group than the Cornell-Coxe when compared to the Binet.) The experimenters suggested that the norms for mental defectives were not adequate and that new norms were needed.

1951, Robert T. Glenn (8). Glenn did a study concerned with relating the Leiter to some other reliable measure of intelligence. The other measure he chose was the 1937 revision of the Binet. He also attempted to find a line of best fit, a regression line, whereby Leiter IQ's could be converted into Binet IQ's (8, p.13).

Since this study was made, Dr. Leiter has changed his definition of what the Leiter measures. In a letter to the writer, dated June 13, 1958, he stated, "While the L.I.P.S. is a non-language test, it is not a non-verbal test. The L.I.P.S. appears to be a performance test, but it is not a performance test in the usual sense. Its uniqueness lies in the fact that it makes verbal test material movable, therefore it looks like a conventional performance scale but this likeness is only in the physical makeup of the test and not in its content. . . . The L.I.P.S. has nothing in common with the conventional performance tests such as form boards, object assemblies, knox cubes, and the other sub tests of the Pintner-Paterson scale . . . the L.I.P.S. is . . . really a non-language Binet scale."
The subjects for Glenn's study were students who had been given the Binet at a C.A. of six. An attempt was made to get a normal distribution of scores, but there was a shortage of cases in the normal and higher ranges. The group was picked from schools containing a normal socio-economic background, tending to make the population approximate the standardization group of both tests.

Glenn criticized Leiter's assumption that his test was an adequate substitute for the Binet because so few studies have been made comparing the results of the two tests.

The results of Glenn's study indicate that he obtained a correlation coefficient of .77 between the Leiter and the Binet. This correlation is significant at the 1 per cent level. The regression equations indicated that for a Leiter score of 100 the Binet score would be 101 ± 10.43.

On the specific items of the test, he felt that on VI-2 (Pattern Completion) chance factors may be more important than intelligence. On VIII-1, the shades of gray are not the same on the blocks as on the strip. He and several colleagues disagreed on the correct placement. On XII-1 (Block design) chance completion of the first design may spur efforts but chance failure may block them. Item XII-3 (Facial expressions) he found to be of questionable validity. A consensus of opinion was needed in order to decide which was the correct response.
Glenn's conclusion about the Leiter is that it does not differentiate between normal and above normal intelligence but may more safely be used to differentiate those children who are mentally deficient from those who are of normal intelligence.

Glenn's statement that the Leiter was not an adequate substitute for the Binet would have been more properly based on the type of studies done and their findings, rather than on the number of studies. He also has not proven in his study that the test will differentiate between mentally deficient and the normal children.

TABLE III
THE BINET-LEITER CORRELATION*

<table>
<thead>
<tr>
<th></th>
<th>Binet</th>
<th>Leiter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>93.36</td>
<td>87.70</td>
</tr>
<tr>
<td>S.E.</td>
<td>2.28</td>
<td>2.74</td>
</tr>
<tr>
<td>S.D.</td>
<td>16.59</td>
<td>19.96</td>
</tr>
</tbody>
</table>

\[ r = .77 \pm .05 \text{ corrected } .778 \pm .053 \]

*Summary of test results reported in Robert Glenn's study (8, p. 20).
1950, Trent E. Bessant. In this study by Bessant (2), twenty cases of chronological ages 4 to 58 years were tested on the Binet and the Leiter. The mean IQ's for both groups were about 70, and the standard deviation for the Binet was 22. The correlation between the mental ages of the subjects, as measured by both tests, was .93. The great variability of the chronological ages probably caused the high correlation. Six of the twenty cases had the same C.A. (nine years) and the Spearman rho of these cases was .99. The small size of the sample and the age variability limit generalization possibilities from this particular study.

1950, Guy L. Bond and Leo C. Far. Bond and Far (5) tested a sample of fifty 4th, 5th, and 6th grade students in Minneapolis schools, using the Stanford-Binet and the Gates Silent Reading Test. The reading age of the students was obtained and the children whose reading age (as measured by the Gates test) exceeded their mental age (as measured by the Binet) were defined as readers. Those whose reading age was below their mental age were defined as non-readers. The groups were then matched on the basis of mental age and the percentage of passes for each item of the Binet scale was computed.

One conclusion drawn from this study was that,
The comparative performance of good and poor readers on verbal items indicates that reading ability tends to distort the M.A. obtained on the Stanford-Binet scale by favoring the child with superior reading ability and penalizing the child with poor reading ability... [this] indicates that a need still exists for an individual intelligence test based upon items that do not depend upon a knowledge of the definition of words, upon the ability to read, and upon the manipulation of words in sentences. (5, p. 479)

This was a small study, therefore the possibilities of generalization from it are limited. However, the study seems to have been conducted in such a way that the results can be accepted, at least for the group tests.

1949, Jack Mathews. Mathews (12) proposes the Leiter as a test that would be adequate for the testing of children who have verbal handicaps, are deaf, have speech defects, or cerebral palsy. The criteria he sets forth for a good test for the verbally handicapped are (1) no verbal responses, (2) no verbal instructions or complicated pantomime, (3) relatively free from time limits, and (4) little penalty for inability to share in cultural experiences. He feels that the Leiter adequately meets all these requirements. However, information on the validity and reliability of the test for handicapped and non-handicapped children is inadequate. The items requiring ability to distinguish color and patterns may
may not be useful for some types of handicapped. Children with certain types of brain damage or vision impairment would not be able to respond to differences in color or pattern.

Point number four in Mathew's criteria has immediate bearing upon this study. The non-reader is not able to share in the aspects of his culture that demand reading ability. Furthermore, the quality of his responses would be impaired by an inadequate vocabulary.

1946, Natalie Darcy. A more recent study giving information on the Binet content was made by Natalie Darcy. In this study Miss Darcy attempted to determine the effect of bilingualism upon performance on verbal and non-verbal tests. Using two matched samples of 106 children, one bilingual and one monolingual, she tested the subjects with the 1937 Revision of the Stanford-Binet and the Atkins Object-Fitting Test, Form A.

Two conclusions Miss Darcy drew from her study were that,

1. There were significant differences between the mean IQ's achieved by monolingual and bilingual subjects on the Stanford-Binet scale. These differences were consistently in favor of the monoglots when divided according to age and sex and also when the age groups and sex groups were combined.

2. Since the monolingual and bilingual subjects of the investigation were closely matched as to number, sex, socio-economic status, and age with six months
intervals, and since the performance of the bilin-
gual subjects was significantly inferior to that of
the monolingual subjects on the Stanford-Binet
scale, but significantly superior to the perform-
ance of monolingual subjects on the Atkins Test,
it may be concluded that the bilingual subjects in
this investigation suffered from a language handi-
cap in their performance on the Stanford-Binet.
(7, pp. 40-41)

This would tend to confirm other statements made con-
cerning the heavily verbal aspect of the Binet.

1946, Charlotte Madeley. A second study of the Leiter,
using the 1938 scale, was made by Madeley (11). Her subjects
were 156 high school students at Placer High School, Auburn,
California. The group was an "unselected group of middle
class children of Nordic stock" (11, p. 97).

The ages of the sample cases varied from 14 years, 0
months to 17 years, 11 months. The S.D. of the scores ranged
from 11.55 to 14.05 points. The group S.D. (on the Leiter)
was 12.00. The mean IQ was 101.84. This mean IQ is almost
7 points above the norm for the Leiter (95).

Madeley suggested that odd year tests be eliminated,
the stronger odd year tests substituted for the weaker even
year tests, credit be doubled for the tests to make up for
the loss of credit from the eliminated tests, and a time
limit be placed on the upper end of the scale (11, p. 111).
She found that the test took far too long to administer --as
much as four hours in one case.
In 1941 Ruth Williams (14) made what she called a "clinical" study of the 1940 revision of the Leiter, using the Leiter and Binet together to determine the relationship between them. The sample was composed of fifty middle class American white children from the ages of 6 years to 10 years. The sample was chosen on the basis of previous test scores (on an unnamed test). The mean score for her group on this test was 108.5 with a S.D. of 11.35.

Williams found that the 1940 Leiter did not show a significantly different correlation with the group test than the Binet. The Binet and Leiter together correlated .67. Using a mental age score rather than an IQ score, she found the correlations for the two tests increased to .888.

In comparing IQ scores, five cases were lost in the six-year group because the children were selected on the basis of teaching ratings rather than group tests.

This study seems to have been an attempt to validate the Leiter scale. The results of the study do not seem to indicate that a relationship between the Leiter and the Binet has been demonstrated, a relationship of a significantly higher order to justify the interchangeability of the tests.
TABLE IV

THE INTERCORRELATIONS OF THE RELATIONSHIP OF
THE LEITER TO THE STANFORD-BINET

<table>
<thead>
<tr>
<th></th>
<th>Group test</th>
<th>Binet M.A.</th>
<th>C.A.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IQ Score</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1940 scale</td>
<td>.45</td>
<td>.67</td>
<td></td>
</tr>
<tr>
<td>Group tests</td>
<td>x</td>
<td>.43</td>
<td></td>
</tr>
<tr>
<td><strong>MA Score</strong></td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1940 scale M.A.</td>
<td>.888</td>
<td>.822</td>
<td></td>
</tr>
<tr>
<td>Binet M.A.</td>
<td>x</td>
<td>.817</td>
<td></td>
</tr>
</tbody>
</table>

*Based on findings in Ruth William's study.
1939, Morris Krugman. Krugman (9) made a study comparing the Binet scale scores with those of non-verbal tests and studying the construction of the Binet. One of the assumptions made by the author in this thesis is that the Binet scale is weighted toward verbal ability. Terman, in Measuring Intelligence, found "that it is extremely difficult to devise non-verbal tests for the upper levels which satisfy the requirements of validity, reliability, and time economy" (13, p. 5).

After the revised Binet scale appeared, Dr. Krugman and his associates at the Bureau of Child Guidance in New York City studied the revision. He used four divisions in his study:

1. Impressions of the test by ten of the staff psychologists.

2. Examination of twelve hundred cases to check the basal and ceiling years.

3. Comparison of the old and new form using a sample of ninety clinical cases.

4. A test of four hundred randomly chosen school children, using the new Form L.

The result of this study was that they found several administrative shortcomings in the Binet, plus an overemphasis on verbal material, especially in the middle and upper ranges (9, p. 602).
IV. SUMMARY

In this chapter the development of both the Leiter and the Binet, and the results of pertinent studies using these instruments, were discussed. The studies seemed to indicate that the Leiter needs more adequate norms for some groups and that no definitive study has been made that will allow the statement to be made that the Leiter is a "non-language Binet."

Two studies illustrating the verbal weight of the Binet are included since one hypothesis of this thesis is that the Binet will tend to penalize persons who are weak in verbal areas.
CHAPTER III

THE ASSUMPTIONS, HYPOTHESIS, AND PROCEDURE

In this chapter the hypothesis, experimental design and procedure of the study will be discussed. The definition of the sample, procedure for administering the Leiter scale, and the statistical techniques used to check the results of the study will be included.

I. THE ASSUMPTIONS

Before making a specific hypothesis the following assumptions should be stated:

1. If readers and non-readers are matched on Binet IQ's, and if the Binet is weighted toward verbal factors, then the non-reader's IQ scores may not be validly assessed.

2. The Leiter is a non-verbal test whose IQ assessment is not artificially lowered by the test instrument and is thus equally valid for readers and non-readers.

II. THE HYPOTHESIS

This study is based on the hypothesis that, using a test of the significance of the difference between two related measures (15, p. 226), there should be a significant
difference between the Leiter IQ's for the readers and non-readers matched on Binet IQ's, with the non-readers receiving higher scores.

The sample included twenty-four elementary school students, ages 9 years, 6 months to 11 years, 9 months. The entire sample had been administered the Binet by the North Sacramento Unified School District School Psychologist. All of the subjects in the sample had been referred to the School Psychologist for testing, therefore it cannot be assumed that they represent typical school children. The reasons for referral of these children were not considered in the selection of the reading sample; if reading difficulty was specified in the referral, the child was placed in the non-reading group, provided all other criteria were satisfied.

No attempt was made to control for sex, race, or income group. The average annual income of this area is below that of many districts in this section of California. The results of this study may be applicable only to similar groups of children.

All of the subjects in the sample had Binet IQ scores between 90 and 125. It was felt that those with scores below 90 would have reading difficulty because of intellectual factors. The children were divided into two groups, matched on the basis of age and intelligence. Group A consisted of
twelve students who were not referred because of reading difficulties. It was determined that they read at least as well as other members of their age group at the time of referral. The actual reading grade placement was checked at the time of testing. Group B contained students whose reading was more than one grade below their class level at the time of the initial referral. Group B was then defined as "non-readers." All of the subjects were assigned to Group A or Group B on the basis of reading and Binet scores. The cases were arranged so that every person in Group A was matched within six points of IQ with a member of Group B. The difference in scores ranged from 0 to 6 points, with the mean difference 3.2 points. The modal difference was 3.0 points. The mean of the reading group was 106.8 and the mean of the non-readers was 104.8. There is no significant difference between the IQ scores for the two groups.

III. THE PROCEDURE

The Leiter scale was administered to the children from both groups in random order -- all children from the same school being tested on the same or adjacent days regardless of IQ, or group placement. The test was administered in accordance with the instructions in the Leiter test manual.
The only change made was that instead of mentioning the "deaf people" aspect (10, p. 264), the children were told, "I am going to ask you to do some things for me; the only rule is that you can't ask me how to do them -- you must figure that out for yourself." This change was made because the first few children tested reacted to the original instructions by not speaking to the tester at all. It was felt that it would be better for rapport and gaining information about the test if the children were encouraged to feel free to speak. Actually, they did not talk very much once the test was in progress. Whenever a question was asked as to the proper placement of a block the response was, "Well, put it where you think it should go" (10, p. 265).

The test was started at age five for all subjects. After testing the first four subjects it was found that the test could not be started above year six because of the sequence of items. Certain items are dependent upon the prior administration of others. The test continued until all items were missed for two consecutive years. The average time for testing was approximately one hour and fifteen minutes. One test took two hours to administer. Most of the tests were given in the afternoon, either at 1:00 or 2:30 p.m. Morning tests were given at 9:00 and 10:30 a.m. The tests were administered in a quiet room -- usually the nurse's office or
the teachers' lunch room. Since the children had all been tested recently by the School Psychologist, they had some idea of what the test situation would be like and it was not necessary to spend much time explaining the process. Most of the children commented when they came in, "I know about this, I did it before." The usual response to this remark was, "This is a little different."

The null hypothesis tested was that there was no significant difference between the Leiter IQ's for the reading and non-reading group.

Using the formula previously reported, the mean difference was 2.25, the standard error of the difference was 5.46, and the resulting t was .41. With a sample N of 12, this difference was not significant at or above the .75 level of significance. Therefore the null hypothesis could not be rejected.

Using the test that independent populations have the same correlations (16, p. 256), the result was that z = .59. With so small a value of z it would not be possible to assume that the relationship is higher for readers than non-readers. Thus non-readers did not improve significantly on the Leiter.

The scores and correlations for the two groups are presented in Table V.
# Table V
**Reader and Non-Reader IQ Scores Based on the Leiter and Binet**

<table>
<thead>
<tr>
<th></th>
<th><strong>IQ Scores</strong></th>
<th></th>
<th><strong>IQ Scores</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Readers</strong></td>
<td><strong>Non-Readers</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Subjects</strong></td>
<td><strong>Binet</strong></td>
<td><strong>Leiter</strong></td>
</tr>
<tr>
<td>1</td>
<td>93</td>
<td>102</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>94</td>
<td>93</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>98</td>
<td>96</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>101</td>
<td>103</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>102</td>
<td>88</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>102</td>
<td>93</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>102</td>
<td>112</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>108</td>
<td>106</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>115</td>
<td>98</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>119</td>
<td>96</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>123</td>
<td>89</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>125</td>
<td>90</td>
<td>12</td>
</tr>
</tbody>
</table>

$\overline{z} = 1282 \quad \bar{z} = 1166 \quad \overline{z} = 1258 \quad \bar{z} = 1183$

$\bar{z} = 106.8 \quad \bar{z} = 97.3 \quad \bar{z} = 104.8 \quad \bar{z} = 98.6$
The correlation coefficients between the Leiter and the Binet for Group A - Readers, Group B - Non-readers, and for the total number of subjects are:

<table>
<thead>
<tr>
<th>Correlations</th>
<th>Readers</th>
<th>Non-readers</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 12</td>
<td>r = -.26</td>
<td>r = .52</td>
</tr>
</tbody>
</table>

Total r = .29

The confidence interval for total r equaled -.26 to .70 (z of 2.58), and the Leiter S.D. was 9.8.
CHAPTER IV

CONCLUSIONS, DISCUSSION, AND SUMMARY

I. CONCLUSIONS

The size of the sample in this study was not large enough to allow any broad conclusions concerning the correlations between the Leiter and the Binet. However, the confidence limits indicate that the tests do not correlate as well as Leiter's table indicated (Table I, page 6). It is not probable that the correlation (at least for this type of homogeneous sample) is as high as .81. The test does not seem to show in this study that it is a "non-language Binet."

The Leiter is not difficult to administer or score, but it is not economical either in money or time. It provides very little personal information about the subject, and it makes rapport difficult to maintain.

II. DISCUSSION

The Leiter can be criticized for several shortcomings:

1. The test was not adequately standardized. In comparison with the careful standardization of the 1937 Revision of the Stanford-Binet, the Leiter is weak in this area. Although many small studies have been done with the test, there
32.

has not been the large scale validation that other such in-


tsuments have had.

2. Sequence of items. Because certain items must be

preceded by others, it is very difficult to start the test-
ing above the six-year level. For example, 18-2 (Dot Estima-
tion) must be preceded by 16-3, 9-1, 8-4, 7-3, and 6-1. If

it is possible to start tests in the middle of such a sequence,
it is not mentioned in the test manual.

3. Failure to complete a block design at one level

seems to carry on to the higher level and the constant repe-
tition of block designs that cannot be completed correctly

seemed to disturb the subjects.

4. The eight-item ceiling is very hard on both the
tester and the subject, especially since the test was started
so low. When the items become too difficult and the child is
conscious that the items are being failed, it becomes increas-
ingly difficult to maintain rapport. The test would be im-
proved if the items discriminated better between the age levels
so that failure of four items would suffice to stop testing.

III. SUMMARY

The object of this study was to determine the relation-
ship between scores of readers and non-readers on the two
measures of intelligence, the Leiter International Performance
Scale and the 1937 Revision of the Stanford-Binet Test of General Intelligence. The children who could read normally did not do as well on the Leiter as those who read poorly, but the difference was not significant, possibly due to the small size of the sample. Further work should be done in this area using a larger group of non-readers, in order to ascertain the degree of relationship between the factors measured by the Leiter and the Binet.

If a larger sample had been drawn, and the difference between the correlation coefficients had remained as extreme (readers, -.26; non-readers, .52), it would be possible to assume that the factors present in the Leiter seem to penalize verbal ability rather than reflect it. A test that measures the same capacities and abilities as the Stanford-Binet should not have produced such discrepant correlation coefficients. The individual cases were discrepant. For example, in the reading group subjects 10, 11, and 12 dropped as much as 34 points from their Binet IQ scores to their Leiter IQ scores.

A test with a high coefficient of correlation (+.80) with the Binet should have produced IQ scores that would be in approximately the same rank order as the Binet IQ's. The Leiter scores were not near to the order of the Binet scores. For example, in the non-reading group, subject number one,
with a Binet IQ of 90, received a Leiter IQ of 99 while subject nine, with a Binet IQ of 112 had a Leiter IQ of 85. These varying scores seem to indicate that the Leiter is not measuring the same factors as the Binet or, possibly, that the Leiter is not a valid test if validity is defined as "equal to the Stanford-Binet."

The difficulties encountered in the experiment due to the sequence of the test items and the poor discrimination between year levels indicate that further work should be done in the area of test procedure and discrimination. Some statement clarifying the procedures with regard to the test administration should be included in the administration manual.

This study also indicated that the Leiter seemed to contain non-verbal rather than non-language items. It would seem advisable that this possibility be investigated.


The scale is administered by placing the matching blocks in a notched frame which holds the pasteboard strip containing the test item.