A Study of Certain "Difficult" Words Used in the Administration of the Revised Stanford-Binet Scale, Form L

George Kenneth Zak
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_theses
Part of the Psychology Commons

Recommended Citation
https://ecommons.luc.edu/luc_theses/1338

This Thesis is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Master's Theses by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 1954 George Kenneth Zak
A STUDY OF CERTAIN "DIFFICULT" WORDS USED
IN THE ADMINISTRATION OF THE REVISED
STANFORD–BINET SCALE, FORM L

by

George Kenneth Zak

A Thesis Submitted to the Faculty of the Graduate School
of Loyola University in Partial Fulfillment of
the Requirements for the Degree of
Master of Arts

June
1954
LIFE

George Kenneth Zak was born in Oak Park, Illinois, September 8, 1925.

He was graduated from St. Patrick Academy, Chicago, Illinois, June, 1943. The next two years were spent in military service. After studying a year and a half at De Paul University, Chicago, he transferred to Loyola University, Chicago, and graduated in June of 1949 with the degree of Bachelor of Science in psychology.

From September, 1949 until August, 1951, the author worked as an assistant psychologist at the Loyola Center for Guidance. During this same period he pursued graduate study in psychology at Loyola University.

Since September, 1951, the author has been employed as an assistant industrial psychologist at the Hawthorne Works of the Western Electric Company, Cicero, Illinois.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. REVIEW OF THE LITERATURE</td>
<td>6</td>
</tr>
<tr>
<td>III. PROCEDURE IN SECURING DATA</td>
<td>13</td>
</tr>
<tr>
<td>IV. AN ANALYSIS OF THE FINDINGS</td>
<td>17</td>
</tr>
<tr>
<td>V. SUMMARY AND CONCLUSIONS</td>
<td>34</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>37</td>
</tr>
<tr>
<td>APPENDIX I</td>
<td>38</td>
</tr>
<tr>
<td>APPENDIX II</td>
<td>39</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. REVIEW OF THE LITERATURE</td>
<td>6</td>
</tr>
<tr>
<td>III. PROCEDURE IN SECURING DATA</td>
<td>13</td>
</tr>
<tr>
<td>IV. AN ANALYSIS OF THE FINDINGS</td>
<td>17</td>
</tr>
<tr>
<td>V. SUMMARY AND CONCLUSIONS</td>
<td>34</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>37</td>
</tr>
<tr>
<td>APPENDIX I</td>
<td>38</td>
</tr>
<tr>
<td>APPENDIX II</td>
<td>39</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>CHRONOLOGICAL AND MENTAL AGES OF THE 56 CHILDREN REPRESENTED IN THE PRESENT STUDY</td>
<td>17</td>
</tr>
<tr>
<td>II.</td>
<td>NUMBER AND PER CENT OF SUBJECTS WHO COULD NOT DEFINE WORDS WHICH APPEARED BETWEEN THE BASAL AGE AND THE HIGHEST LEVEL OF SUCCESS</td>
<td>21</td>
</tr>
<tr>
<td>III.</td>
<td>CORRELATION BETWEEN BINET VOCABULARY TEST AND MENTAL AGE FOR THE SUBJECTS USED IN THE PRESENT STUDY AND AS REPORTED BY ELWOOD</td>
<td>23</td>
</tr>
<tr>
<td>IV.</td>
<td>A COMPARISON OF THE SUBJECTS’ WORD DEFINING ABILITY WITH THEIR PERFORMANCE ON THE RELATED BINET ITEMS</td>
<td>27</td>
</tr>
<tr>
<td>V.</td>
<td>NUMBER AND PER CENT OF SUBJECTS BY MENTAL AGE WHO WERE ABLE TO DEFINE EACH &quot;DIFFICULT&quot; WORD</td>
<td>32</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>DISTRIBUTION OF THE IQ SCORES OF THE 56 CHILDREN</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>REPRESENTED IN THE PRESENT STUDY</td>
<td></td>
</tr>
</tbody>
</table>

vi
CHAPTER I

INTRODUCTION

One of the most essential parts of a psychometric examination is a determination of the level of intelligence. This is perhaps especially true in working with children, for many of their problems are related in some way to their mental capacity. A knowledge of a child's intelligence level is a key to estimating his potentiality for success in school, establishing his curriculum at a level and at a rate suitable for him, detecting unusual strengths or weaknesses among his mental capacities, and many other important problems.

The study of the measurement of intelligence, begun more than fifty years ago, has progressed so well that many individuals, both in and out of the field of psychometry, look upon experimental work in this area as a waste of effort. Although this attitude reflects well on the diligence and ingenuity of experimenters such as Cattell, Binet, Simon, and Terman, it is nonetheless true that much work is yet to be done.

The purpose of the present study is to examine in detail certain components of the Revised Stanford-Binet Scale, Form L, regarded today as the most valid test of intelligence.
for children. The present revision is the fifth in the Binet series. The first test in the series, introduced in France in 1905 by Binet and Simon, was the first successful psychometric method of detecting whether a child had sufficient capacity to profit from attendance at an ordinary school. This test was revised in 1908 and again in 1911.

The method was introduced in the United States in 1916 by Terman, who authored the Stanford Revision of the Binet. The Stanford Revision proved to be very valuable in studying children of all mental levels, and soon became the standard by which other intelligence tests were judged. The Revised Stanford-Binet was introduced by Terman and Merrill in 1937 to replace the Stanford Revision of 1916. It was hoped that the more gross shortcomings inherent in the 1916 revision had been overcome, and that the Revised Stanford-Binet would be even more valuable.

One of the more serious shortcomings of the 1916 revision was its heavy reliance on verbal items. The result of this was that children with good reading ability and a well-developed vocabulary were favored by this test while those less accomplished in this respect were penalized.

Since the present revision was introduced it has been subjected to a great deal of analysis. On the whole, this research has substantiated the claims for its high validity. As would be expected, however, clinical experience with this
instrument has led many workers to suspect flaws which were not foreseen in its preparation. The present study is an attempt to explore experimentally an hypothesis formulated through experience with the Revised Stanford-Binet concerning a suspected weakness.

This hypothesis is that certain words, not in the vocabulary test, but used elsewhere in instructions or content, are beyond the comprehension of children at the age level where these words are found. Based on the observations of several experienced psychologists at the Loyola Center for Guidance, the following words were selected for study: thirsty, cane, blunt, bound, creased, trousers, skull, iceberg, Gulf Stream, collision, and proverb. The first word appears in the Binet at the III-6 level, the next two words appear at the IV-6 level, the next at the VIII-year level, the next five appear at the IX-year level, the next at the XI-year level, and the last word appears at the Average Adult level. Six of these words appear in verbal absurdities, two in opposite analogies, one in the directions for the paper cutting test, one in a comprehension test, and one in a proverbs test.

If this hypothesis were proved to be correct it would mean that children are not being given an adequate opportunity to demonstrate their capacity for success on those parts of the scale which contains these words.

The authors of the scale state that intelligence can
best be measured by testing an individual on many different kinds of mental endeavor, and they included a wide variety of tasks in the Revised Stanford-Binet. The vocabulary test, while demonstrated to be the best single item in the scale, is only one of the many avenues the scale explores in establishing an individual's intelligence. Furthermore, the vocabulary level of an individual, while a useful rough index of intelligence, is not in itself the complete answer. Many other tasks within the scale having high correlations with the total-score criterion have little if any connection with the extent of the subject's vocabulary.

If, then, success on parts of the scale other than the vocabulary test depends in whole or in part on a knowledge of the meaning of words, those parts of the scale are not measuring what they purport to measure. In these cases only those individuals who have a very good vocabulary would be in a position to demonstrate whether or not they can do the task in a successful manner. Those individuals whose knowledge of words is rather limited for one reason or another would have little or no chance of scoring a success on a test item which supposedly is not measuring vocabulary.

The purpose of the present study, then, is to determine whether or not the inclusion of certain "difficult" words constitutes a weakness in the scale. If the present evidence sustains this suspicion, it might promote the study of a much
neglected aspect of test construction. It would indicate that the authors of tests, to insure a higher validity, would have to select the words used in the instructions to the subject or in the content of test items on a more sound basis than now employed. Seldom is any reference made to this problem in the presentation of the techniques and controls used in the construction and standardization of a psychological test. It is hoped that the present study will shed some light on this problem, and perhaps stimulate further advance along a little-used approach to increased test validity.
CHAPTER II

REVIEW OF THE LITERATURE

A review of the literature available on the Revised Stanford-Binet is no small undertaking. Research with this instrument has been carried on with considerable enthusiasm beginning with the Binet-Simon scales, and in recent years the psychological and educational journals have reported a wealth of findings concerning this test. In fact, approximately five hundred research studies were found in the literature which relate to the Revised Stanford-Binet.

A great deal of the literature is concerned with research on limitations connected with various verbal aspects of the Binet, but very few deal directly with the present problem. Most of this research pertains to such problems as the effects of bi-lingual homes, the relative difficulty of test items, and the performance of poor readers.

Before presenting findings which are directly related to the present problem, it would be well to review some of the research on verbal limitations as a whole. To begin, the authors of the Revised Stanford-Binet, in discussing its essential features, clearly show they are aware of the fact that its
too verbal character is a limitation. Although they attempted to increase the number of non-verbal tests throughout the scale, they claim real success only at the lower levels. However, they defend the use of preponderantly verbal tests at the upper levels on grounds that at this level intelligence is largely a matter of ability to do conceptual thinking.¹

In discussing non-intellectual elements which might affect the score on this scale Cronbach² sums up the results of research by saying that the Revised Stanford-Binet underestimates the intelligence of "bilingual children, children from homes where experiences with English are limited, children with hearing deficiencies, and children who have difficulty in reading." He is not, of course, taking exception to the well-known fact that verbal tests are the most efficient means of measuring intelligence, but merely that this approach is not valid under certain conditions.

One of the earliest studies of verbal limitations of the current revision was published in 1938 by E. A. Bond³ who studied one hundred and thirty good and poor readers in the

---

¹ Lewis M. Terman and Maud A. Merrill, Measuring Intelligence, New York, 1937, 4-5.


ninth grade who were matched on the basis of sex, CA and IQ. He found that the poor readers were inferior on vocabulary, abstract words, and sentence building. He then regraded each of the Binet tests, omitting the verbal items. As a result of this, some of the IQ scores were raised as much as fifteen points, and the average IQ for the poor readers was higher than the average IQ for the good readers. He concludes that these verbal items should be omitted in testing poor readers on the Revised Stanford-Binet.

A study which was almost a duplicate of E. A. Bond's work was published in 1950 by Guy L. Bond and Leo C. Fay. In this study they used as subjects one hundred children from the fourth, fifth, and sixth grades of a Minneapolis elementary school. A composite reading age was determined on the basis of four Gates tests, and a child was classified as a poor reader if his reading age was less than his mental age.

A comparison of the percentages of each group passing each item revealed statistically significant differences between the good and poor reader groups. They conclude that poor readers are definitely handicapped by those items which are verbal in nature. Further, the greater the reading disability,

the more pronounced this handicap seems to be. Bond and Fay summarize their findings as follows:

The comparative performance of good and poor readers on verbal items indicates that reading ability tends to distort the mental age obtained on the Stanford-Binet Scale by favoring the child with superior reading ability and by penalizing the child with poor reading ability. This tendency, which is that observed with group intelligence tests that are verbal in nature, 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

It is worth noting that both studies made on good and poor readers conclude in part that a knowledge of the definition of words is a definite advantage to the subject. It is not clear, however, from the texts of the published reports, whether they are objecting to the vocabulary test in the Revised Stanford-Binet or to words found elsewhere in test items or instructions. It appears at least probable to the present writer that these studies referred to items where the words themselves were not intended to be discriminative. If this is true, then these studies seem to bear more directly on the present problem. The basis for this conclusion is the well-established fact that a vocabulary test is an excellent means of measuring intelligence if the individual has been exposed to a reasonably normal environment. The authors of the Binet judge the vocabulary test to be the best one in the entire

5 Ibid., 477.
scale. They report correlations between the vocabulary test and mental age for single age groups ranging from .65 to .91 with an average of .81. Elwood found a correlation of .978 between the vocabulary test and mental age for a group of 1,161 Pittsburg school children tested during one year.

Additional insight into factors related to success or failure in grasping the meaning of words was provided in 1936 by Feinberg in a study using the 1916 Stanford Revision. In a study using 1,431 children and adults tested at a mental hygiene clinic as subjects, he found that the ability to define the word "mellow" is dependent not alone on intelligence but also on the age and education of the subject. The present study is an attempt to study whether certain words included in the scale are, for this reason, unsuitable at the levels on which they occur.

Perhaps the most comprehensive report related to the present study was made by Krugman in 1939. As part of a rather

6 Terman and Merrill, *Measuring Intelligence*, 302.


8 Henry Feinberg, "The Examinee Defines 'Mellow,'" *Journal of Educational Psychology*, XXVII, 1936, 191.

thorough examination of the Revised Stanford-Binet Scale, he collected the opinions of ten psychologists from the Bureau of Child Guidance of New York City who together had given approximately five thousand Revised Stanford-Binet Scales.

He found high agreement among these ten psychologists, from whose comments he drew many conclusions. The conclusions having a direct bearing on the present study were that the emphasis on verbal material is still present, especially in the middle and upper levels, that years VIII and XI are the levels most dependent on verbal material, and that many words seem unfair.\(^{10}\)

Specific examples related to each of his conclusions are listed at length. Included under items frequently mis-interpreted were all the verbal absurdities, mainly because of language difficulty, and the paper cutting test. Among items considered unfair to New York City children were verbal absurdities and opposite analogies, again because of a dependence on language.\(^{11}\) Of the eleven words included in the present study, six appear in verbal absurdities, two in opposite analogies, and one in the directions for the paper cutting test.

There can be no doubt that verbal limitations exist

\(^{10}\) Ibid., 602.

\(^{11}\) Ibid., 596.
regardless of the particular concepts held on the nature of intelligence and on the most expeditious methods of measuring it. A survey of the literature indicates that the present problem has been recognized in a general way, but no experimental studies of these or similar words have been reported.
CHAPTER III

PROCEDURE IN SECURING DATA

In order to test the hypothesis that certain words in the scale are beyond the comprehension of the children being tested on the Revised Stanford-Binet, it was decided that subjects would be secured both from the Loyola Center for Guidance and from local grade schools. The subjects tested at the Loyola Center for Guidance would in the main be children who were referred for a variety of behavior and school adjustment problems. It was not the intention to make a comparative study between these two groups, but merely to get more of a cross section of children in general.

So that any conclusions might be as specific as possible, the subjects selected for this study were those whose chronological ages were at least five years and below twelve years, and whose IQ's were 80 or above. Setting these ranges also had the effect of eliminating from the study children whose mental ages fell below four years.

Twelve examiners participated in gathering the data for this study. Four experienced psychologists and four graduate interns at the Loyola Center for Guidance followed the
procedures outlined below as a routine during an eight-month period, including all children who fell within the proper chronological and IQ ranges. During this same period four graduate students who had administered at least 25 Revised Stanford-Binet Scales under supervision followed the same procedure in testing school children in local grade schools. Slightly more than half of the subjects were tested at the Loyola Center for Guidance.

It was of course necessary to devise a method of measuring the amount of comprehension children have for each of the key words. It was decided that a supplementary oral vocabulary test modeled after the Binet vocabulary test would be eminently suited to the present problem. And in order that a child's experience on the supplementary vocabulary test would have no effect on his performance on the Binet itself, each child was given the Binet before he was given the supplementary vocabulary test.

The procedures and standards for administering and scoring this supplementary vocabulary test were patterned after the techniques and criteria used for the Binet vocabulary test, with one important addition. If the child was unable to give a satisfactory definition of the word, the word was then presented in a sentence to ascertain whether hearing it in context would permit the child to give evidence of at least a lesser level of understanding. See Appendix I for a copy of the directions used in giving the test. In order to maintain complete uniformity,
the sentences containing each of the key words were selected in advance. Care was taken to see that the sentences chosen were not similar to Binet material in content. A verbatim account of each response was kept on a record blank. After the Binet was scored a record of the successes and failures on items containing the key words were transferred to the record blank. Other information such as date, sex, CA, MA, IQ, score on Binet vocabulary test, basal age, and the highest level of success on the Binet was also recorded. See Appendix II for a sample of the record blank.

As would be expected, some of the responses given in the supplementary oral vocabulary test were difficult to score because of a lack of clarity or precision in verbal expression on the part of the subjects. This same difficulty is inherent in the Binet itself, and accounts in part for the fact that a competent Binet examiner must have a high degree of judgment, intelligence, and training.

Since no specific precedent for scoring these responses exists in the Binet scoring standards, it was decided that the judgment of experienced psychologists would be utilized in all questionable cases. It was found that somewhat less than thirteen per cent of the responses on the supplementary oral vocabulary test fell into this category. These questionable responses, therefore, were scored independently by five psychologists who were thoroughly familiar with the Binet, and the
final score assigned each questionable response represented the majority opinion. The fact that all five judges agreed on the scoring of 62.5 per cent of the questionable responses, and at least four agreed on the scoring of 84.1 per cent of the questionable responses would seem to indicate that the oral vocabulary test was both reliable and valid.

In all, seventy subjects who fell within the chronological age range of five years to eleven years and eleven months were tested. However, thirteen had to be dropped because their IQ's fell below 80, and one had to be dropped because his performance was marked by pronounced psychotic tendencies.
CHAPTER IV

AN ANALYSIS OF THE FINDINGS

Before presenting the data the characteristics of the subjects used in this study should be analyzed. Table I shows the distribution of the chronological and mental ages as well as the frequency of each combination. As mentioned in Chapter III, the subjects selected for this study were those whose

TABLE I

CHRONOLOGICAL AND MENTAL AGES OF THE 56 CHILDREN REPRESENTED IN THE PRESENT STUDY

<table>
<thead>
<tr>
<th>C. A.</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>7</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td></td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>12</td>
<td>10</td>
<td>6</td>
<td>9</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

17
chronological ages were above five years and below twelve years, and whose IQ's were 80 or above. These limits would also elimi-
inate any children whose mental ages fell below four years. As it turned out, the mental ages of the group range from five years-five months to thirteen years-two months. Eighty per cent of the children studied range from seven through eleven years in mental age. About 39 per cent of the children have a chronol-
ogical age between eight and nine, and about 43 per cent have a mental age at these same levels.

Figure 1 shows the distribution of the IQ scores for

![Histogram showing distribution of IQ scores]

FIGURE 1

DISTRIBUTION OF THE IQ SCORES OF THE 56 CHILDREN REPRESENTED IN THE PRESENT STUDY
the 56 subjects. They range from 80 to 135. The distribution is negatively skewed, primarily because of the fact that subjects who scored below 80 were eliminated from the study. While there are 29 subjects who scored above the normal range of 90 to 109, there are only seven subjects who scored below it, that is, between 80 and 89. Another factor which contributed to the skewness is the unusually large percentage of superior and very superior children found in the sample. The mean of the IQ scores is 109.3 (S.D.=14.31), and the median is 110.2 (Q=10.0).

Now if the scale was designed so that success on items other than the vocabulary test items depends on such mental abilities as the capacity to detect absurdities or draw analogies, then the words used in the instructions for, or the content of the test items should be easily understood by the children being tested at that particular level. If the words were not understood by these children, then the test item might be failed, not for a lack of the ability to detect an absurdity, for instance, but because an essential part of the problem was not presented in a clear manner.

It clearly follows, then, that no non-vocabulary test item on which a child's mental age score is based should contain words which are beyond that child's level of understanding.

The mental age level is derived on the Binet by finding first a mental age level on the scale where the child passes all items, then testing on successively higher levels until the
child fails all items on a mental-age level. The assumption that failure on an item reflects an inability to solve a problem of a certain type and level of difficulty would of course have to hold true throughout the entire range of items given. It is reasonable to demand, then, that all non-vocabulary words appearing throughout this range be understood by the subjects being tested.

An examination of Table II shows that this condition was not satisfied for almost all of the "difficult" words studied. The most striking results were obtained for the words "Gulf Stream," "blunt," and "proverb" which could not be defined in 94 per cent or more of the 92 occasions that these words occurred in the Binet with various subjects, even after they heard the words in context. Yet in the Binet Scale these same children were asked to solve problems in which knowledge of these words might have been a prerequisite.

Next in order are found the words "bound," "skull," and "iceberg," for which the percentages of failure were 68.7, 57.4, and 55.3 respectively. And the next two words in order of importance are "collision" and "creased" for which the percentages of failure were 36.6 and 31.9 respectively.

The word "thirsty" appears on the III-6 level, and, as it turned out, a basal age was established above this level in all 56 cases. For this reason the level of understanding the subjects have of this particular word has no bearing on this
TABLE II
NUMBER AND PER CENT OF SUBJECTS WHO COULD NOT DEFINE WORDS WHICH APPEARED BETWEEN THE BASAL AGE AND THE HIGHEST LEVEL OF SUCCESS

<table>
<thead>
<tr>
<th>Word</th>
<th>Number of Cases</th>
<th>Number who could not define word</th>
<th>Per cent who could not define word</th>
</tr>
</thead>
<tbody>
<tr>
<td>thirsty</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>cane</td>
<td>38</td>
<td>5</td>
<td>13.2</td>
</tr>
<tr>
<td>blunt</td>
<td>38</td>
<td>36</td>
<td>94.7</td>
</tr>
<tr>
<td>bound</td>
<td>48</td>
<td>33</td>
<td>68.7</td>
</tr>
<tr>
<td>creased</td>
<td>47</td>
<td>15</td>
<td>31.9</td>
</tr>
<tr>
<td>trousers</td>
<td>47</td>
<td>4</td>
<td>8.5</td>
</tr>
<tr>
<td>skull</td>
<td>47</td>
<td>27</td>
<td>57.4</td>
</tr>
<tr>
<td>iceberg</td>
<td>47</td>
<td>26</td>
<td>55.3</td>
</tr>
<tr>
<td>Gulf Stream</td>
<td>47</td>
<td>46</td>
<td>97.9</td>
</tr>
<tr>
<td>collision</td>
<td>41</td>
<td>15</td>
<td>36.6</td>
</tr>
<tr>
<td>proverb</td>
<td>7</td>
<td>7</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Totals      | 407             | 214                              | 52.6                              |

particular aspect of the investigation.

Table II also shows, in a summary way, that in 52.6 per cent of the occasions that these words appeared between the basal age and the highest level of success the subjects were unable to define them. The importance of these findings rests,
of course, on the assumption that the inability to define a word is an indication that the subject does not understand the word well enough to make any use of it.

The basis for this assumption can be found in the Binet Scale itself. The supplementary vocabulary test used in this study was carefully modeled after the Binet vocabulary test both in presentation and the rationale of the scoring. Terman and Merrill\(^1\) state that the purpose of their vocabulary test "is to determine whether the subject knows the meaning of the word . . . ." The procedure used in the presentation of the "difficult" words goes, in fact, even further than the method used in the Binet in attempting to establish the subject's level of understanding. As explained in Chapter III, if the child was unable to give a satisfactory definition of the word, the word was then presented in a sentence to ascertain whether hearing it in context would permit the child to give evidence of at least a lesser level of understanding. And the number and per cent reported as failing to define a word include only those who could define it neither in isolation nor in context.

A more minor objection might possibly be raised on the grounds that this particular group of children might have been unusually deficient in verbal ability. This, of course, would be reflected in their performance on the Binet vocabulary test.

\(^1\) Terman and Merrill, *Measuring Intelligence*, 303.
As indicated in Chapter II, Terman and Merrill reported correlations between the vocabulary test and mental age for single age groups ranging from .65 to .91 with an average of .81. Also reported in Chapter II was Elwood's finding of a correlation of .978 between the vocabulary test and mental age for a group of 1,161 school children of various ages. Elwood's extremely high correlation is due, of course, to the fact that the age factor

was not held constant. The range of ages would in itself cause some positive correlation.

In order to clarify this question, the correlation between the Binet vocabulary test and mental age was computed for the group used in the present study. Since the group includes a wide age range similar to Elwood's group, the obtained correlation must be interpreted only in terms of Elwood's finding. Table III presents the findings of this computation along with the correlation reported by Elwood. As Table III indicates, they are almost identical. It could hardly be maintained, therefore, that the group of children used in the present study was unusually deficient in verbal ability.

Returning to the findings once again, the next step might be an examination of the possible effect of a lack of understanding of these words on the Binet test itself. In many cases some understanding of the meaning of the "difficult" word is a necessary prerequisite to solving the problem. In other cases a lack of understanding of the "difficult" word would not be critical.

An examination of the Binet items in which the various "difficult" words appear would indicate that some knowledge of the words "thirsty," "cane," "blunt," "trousers," "skull," and "collision" is essential in solving the problem. Table IV presents a comparison of the subjects' word defining ability with their performance on the related test items on the Binet Scale.
In order to see if there was any relationship between
the two variables a chi-square was computed for each of the
"difficult" words. The obtained chi-squares indicate with what
degree of confidence we can reject the hypothesis that there is
no significant relationship between knowledge of a word and suc-
cess on the related test item.

Among the six words for which some understanding ap-
ppears to be essential for success on the related test items,
three of them are of especial interest. For the word "skull"
the chi-square was easily significant at the .01 level of confi-
dence. This indicates that there is a definite relationship
between a knowledge of the word "skull" and success on the re-
lated item. And, as Table II indicates, since 57.4 per cent of
the subjects could not define this word, they are handicapped in
a way which was certainly not intended.

The word "skull" appears in a test of verbal absurdities
where the purpose of the test "is to discover whether the
subject can point out the intellectually irreconcilable elements
of the situation presented." 3

The test item is as follows:

In an old graveyard in Spain they have discovered
a small skull which they believe to be that of
Christopher Columbus when he was about ten years
old. What is foolish about that?

3 Terman and Merrill, Measuring Intelligence, 235.
It can be seen that the word "skull" is a crucial one, and any difficulty in understanding it would be an unwanted hindrance to discovering the intellectually irreconcilable element involved in the statement.

A significant relationship between knowledge of the word and success on the related item on the Binet Scale was not quite established for two other "critical" words--"cane" and "collision." The relationship just missed the .05 level of confidence for the word "cane," and fell well within the .10 level of confidence for the word "collision." The chi-squares did not approach significance for any of the other "difficult" words in the study.

Looking at Table IV again, it can be seen that 29 out of 31, or 93.5 per cent of the subjects who passed the related test item could not define the word "blunt." The word "blunt" occurs as one of the key words in a so-called test of "opposite analogies," which appears in the Binet as follows:

Procedure: Say, "The point of a cane is blunt; the point of a knife is . . . ."

In describing this test, the authors state that they have combined the opposites test with the well-known analogies test, in which a fourth word is related to a third as a second is to a first. In this test the clue to the relationship between the second and the fourth of the critical words is that of opposites.  

---

4 Ibid., 216.
Since the second of the "critical" words was not understood by 93.5 per cent of the subjects, this test item, for them at least, was not an opposite analogies test at all. It appears that this test item is passed by a direct association of the idea "sharp" with the idea "knife."

**TABLE IV**

A COMPARISON OF THE SUBJECTS' WORD DEFINING ABILITY WITH THEIR PERFORMANCE ON THE RELATED BINET ITEMS

<table>
<thead>
<tr>
<th>Word</th>
<th>Number of Cases</th>
<th>Defined Word</th>
<th>Did Not Define Word</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Passed Related Item</td>
<td>Failed Related Item</td>
</tr>
<tr>
<td>thirsty</td>
<td>0</td>
<td>29</td>
<td>4</td>
</tr>
<tr>
<td>cane</td>
<td>38</td>
<td>29</td>
<td>4</td>
</tr>
<tr>
<td>blunt</td>
<td>38</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>bound</td>
<td>48</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>creased</td>
<td>47</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>trousers</td>
<td>47</td>
<td>30</td>
<td>13</td>
</tr>
<tr>
<td>skull</td>
<td>47</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>iceberg</td>
<td>47</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>Gulf Stream</td>
<td>47</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>collision</td>
<td>41</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>proverb</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
The data also offer some suggestions as to the relationship between mental age and the ability to understand and define the "difficult" words, although this is not within the scope and design of the present study. In order to study the problem from this point of view, a large random sample would have to be studied at each age level. Since reliability formulas assume a sufficiently large sample, the data at hand on this aspect of the problem cannot be subjected to valid statistical analysis. Garrett\(^1\) states that "If \(N\) is less than about 25, there is usually little reason for assuming such a sample to be descriptive of a given population." The number of subjects at any one mental age level varies from one to twelve in the present study, and therefore does not meet this requirement. Since the subjects at each mental age level are not representative of the general population, any trends in the data cannot be ascribed to the general population. The results are presented only as a guide for further research along these lines.

Table V summarizes the results for the group studied, showing both the number and per cent of subjects by mental age who were able to define each "difficult" word. The results include all 56 subjects, for each subject was asked to define all the words regardless of whether or not they appeared in the Binet

for that individual. In cases where a test item containing one of these words appears in the Binet at two different levels, the word is listed for the purpose of the present study at the lower of the two levels.

Of the eleven words studied, the evidence suggests that at least five of them are too difficult at the level where they appear in the scale.

The two words which appear to be far too difficult for the mental age level at which they appear are "blunt" and "Gulf Stream." The word "blunt" appears at the IV-6 level, but none of the sixteen subjects with a mental age of seven or less was able to define it. One out of twelve with a mental age of eight defined it, one out of ten with a mental age of nine defined it, one out of six with a mental age of ten defined it, and one out of nine with a mental age of eleven defined it. All three subjects whose mental ages were twelve or thirteen failed to define it. Since knowledge of this word is essential in solving the Binet test item in the manner the authors intended, it appears quite clear that the inclusion of the word "blunt" has once again been shown to be less than satisfactory.

"Gulf Stream" appears at the IX-year level. This term was defined by only one of the 56 subjects. This subject had a mental age of thirteen, the highest in the group. Although a knowledge of the meaning of this word is not essential in solving the problem in which it appears, the subject might be misled and
confused by the presence of a word which he does not understand. The evidence at hand, while not conclusive, certainly suggests that the difficulty level of "Gulf Stream" ought to be studied very completely.

Table V also seems to indicate that the words "bound," "skull," and "iceberg" are too difficult at the level at which they appear in the Binet. The word "bound" appears at the VIII-year level, but only one out of twelve who scored a mental age of eight years could define this word. And only thirteen of the twenty-five children who scored a mental age of nine, ten, or eleven could define it.

Somewhat similar results were found for the words "skull" and "iceberg." In both cases less than half of the children whose mental ages coincided with the mental age level at which these words appeared could define these words. And again, a level is never reached where the word is easy for the children. Moreover, as in the case of the word "blunt," some knowledge of the word "skull" is required in solving the test item in which it appears.

The results suggest that three of the words are not too difficult at the level where they appear in the Binet Scale. These words are "creased," "trousers," and "collision." The word "creased" shows a gain in per cent passing from level to level. It appears in the Binet at the IX-year level. Every one of the twelve children whose mental ages were eleven or more was able to
define it, and most of the children with a mental age of nine or ten were able to define it.

The word "trousers," which also appears in the Binet at the IX-year level, is the only word of those studied which gives no trouble at all. The evidence suggests that this word is easy enough to be understood by children with a mental age of about seven.

The word "collision," which appears in the Binet Scale at the XI-year level, was adequately defined by nine out of the ten children who had a mental age of nine, but could not be defined by five of the fifteen children whose mental ages were ten or eleven.

Because of the somewhat limited mental age range of the subjects included in this study, no trends can be suggested for the words "thirsty," "cane," and "proverb." The word "thirsty" appears on the III-6 level, and the minimum mental age of the subjects was five. All 56 subjects were able to define this word, even out of any context. It appears that the word is definitely understood by children with a mental age of five, but nothing is known about this word below this level.

In the case of the word "cane," which appears in the Binet at the IV-6 level, no trends can be discerned because of the small number of children included whose mental ages were near this level.

And finally, the word "proverb," which appears at the
Average Adult level on the Binet, appears beyond the mental age range of the subjects included in the present study. However,

### TABLE V

**NUMBER AND PER CENT OF SUBJECTS BY MENTAL AGE WHO WERE ABLE TO DEFINE EACH “DIFFICULT” WORD**

<table>
<thead>
<tr>
<th>Word</th>
<th>Level in Binet</th>
<th>M. A. of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>thirsty</td>
<td>III-6</td>
<td>No.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>cane</td>
<td>IV-5</td>
<td>No.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>blunt</td>
<td>IV-5</td>
<td>No.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>bound</td>
<td>VIII</td>
<td>No.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>creased</td>
<td>IX</td>
<td>No.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>trousers</td>
<td>IX</td>
<td>No.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>skull</td>
<td>IX</td>
<td>No.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>iceberg</td>
<td>IX</td>
<td>No.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>Gulf Stream</td>
<td>IX</td>
<td>No.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>collision</td>
<td>XI</td>
<td>No.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>proverb</td>
<td>A.A.</td>
<td>No.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
</tr>
</tbody>
</table>
it is interesting to note that none of the subjects was able to define this word, even at the twelve and thirteen year levels.

These observations would suggest, then, that some of these, and possibly other words which appear in the Binet Scale on non-vocabulary test items, are too difficult at the level at which they appear.
CHAPTER V

SUMMARY AND CONCLUSIONS

1. It was suspected that certain words in the Revised Stanford-Binet, Form L, are not understood and cannot be defined by children taking this test. Some of the words appear in the verbal instructions, and others in the content of the test items themselves. In order to test this hypothesis a group of 56 children was given first the Binet and then a supplementary vocabulary test consisting of these suspected "difficult" words.

2. The subjects selected for this study were those whose chronological ages fell between five years and eleven years eleven months, and whose IQ's were 80 or above.

3. The words "Gulf Stream," "blunt," and "proverb" could not be defined in 94 per cent or more of the 92 occasions that these words occurred in the Binet for various subjects, even after the children heard the words in context. The words "bound," "skull," "iceberg," "collision," and "creased" could not be defined in 68.7, 57.4, 55.3, 36.6, and 31.9 per cent of the cases respectfully. In summary, it was found that in 52.6 per cent of the occasions that the eleven words appeared between the basal age and the highest level of success the subjects were unable to de-
fine them.

4. A chi-square test indicated that there is a definite relationship between a knowledge of the meaning of the word "skull" and success on the related test item. This finding was easily significant at the .01 level of confidence.

5. A significant relationship between knowledge of the meaning of the word and success on the related test item was not quite established for the words "cane" and "collision." The relationship just missed the .05 level of confidence for the word "cane," and fell well within the .10 level of confidence for the word "collision." The chi-squares for the remainder of the words did not approach significance.

6. The word "blunt" could not be defined by 93.5 percent of the subjects who passed the related item. Since this word occurs as one of the key words in a so-called test of opposite analogies, this test item, for them at least, was not an opposite analogies test at all. It appears that this test item is passed rather by a process of direct association of the idea "sharp" with the idea "knife."

7. Although it is not within the scope and design of the present study, it was also possible to make some tentative conclusions about the relationship between mental age and the ability to understand and define these "difficult" words. Since the number of subjects at each mental age level is rather small, valid statistical inferences cannot be made about children in
general. However, for the group studied, the words "blunt" and "Gulf Stream" appear to be too difficult for the mental age level at which they appear in the Binet. Also difficult for children with a mental age equivalent to their location in the test were the words "bound," "skull," and "iceberg." An average of only 25 per cent were able to define these three words. Three words, "creased," "trousers," and "collision," do not appear to be too difficult at the level where they appear in the Binet.

8. On the basis of the findings of the present study it appears that a fairly important consideration has been overlooked in the construction of the Binet. Maximum validity is not being achieved because the problems to be solved are made obscure by words both in the instructions to the subject and in the content of the test items which are beyond the comprehension of the children.

9. The results indicate that all words used in tests of this kind should be drawn from lists of words whose difficulty level has already been established.
BIBLIOGRAPHY


Elwood, Mary Isabel, "A Preliminary Note on the Vocabulary Test in the Revised Stanford-Binet Scale, Form L," Journal of Educational Psychology, XXX, 1939, 632-634.


Terman, Lewis M., and Merrill, Maud A., Measuring Intelligence, New York, 1937.
APPENDIX I

PROCEDURE FOR ADMINISTERING THE SUPPLEMENTARY VOCABULARY TEST

AFTER giving the Binet:

1. Say: "I HAVE SOME WORDS HERE WHICH YOU HAVE PROBABLY USED OR HEARD SOMEONE ELSE USE. LISTEN, AND WHEN I SAY A WORD, YOU TELL ME WHAT IT MEANS."

2. Then say: "WHAT DOES......MEAN?" or "TELL ME WHAT A......IS,"
   or "TELL ME WHAT......MEANS," or "WHAT IS A......?"
   or give just the word itself. Record the responses verbatim for each of the eleven words in the spaces provided on the record blank.

3. If the definition that is given is not clear, say:
   "TELL ME WHAT YOU MEAN," or "TELL ME MORE ABOUT IT,"
   and insert (Q) in the text of the responses.

4. If the child is unable to define the word at all, or to explain it well enough, repeat the word, give it in context by reading the sentence on the record blank, and say the word again. Insert (S) in the text of the response to indicate where the sentence was introduced.

5. If the response is still vague, say:
   "TELL ME WHAT YOU MEAN," or "TELL ME MORE ABOUT IT,"
   and insert (Q) in the text of the response.
III-6,6
a) ()

thirsty
I get thirsty on hot days.
cane
The man's cane was crooked.
blunt
It was a blunt end.
bound
They bound the bundles.
crease drawn?
She creased the papers carefully.
trouser
He wore his new trousers to the party.
skull
He cracked his skull when he fell.

iceberg
We saw an iceberg in the ocean.

Gulf Stream
The Gulf Stream starts near Florida.

collision
There was a collision between a bus and a car.

proverb
He often used proverbs to teach a lesson.
APPROVAL SHEET

The thesis submitted by George Kenneth Zak has been read and approved by three members of the Department of Psychology.

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

The thesis is therefore accepted in partial fulfillment of the requirements for the Degree of Master of Arts.

June 12, 1954

[Signature of Adviser]