



1952

Performance of Children of the Same Mental Age on the Revised Stanford-Binet Scale, Form L

Mary Fidelissima Dzik
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_theses

 Part of the [Psychology Commons](#)

Recommended Citation

Dzik, Mary Fidelissima, "Performance of Children of the Same Mental Age on the Revised Stanford-Binet Scale, Form L" (1952). *Master's Theses*. 986.
https://ecommons.luc.edu/luc_theses/986

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 © 1952 Mary Fidelissima Dzik

PERFORMANCE OF CHILDREN OF THE SAME MENTAL AGE
ON THE REVISED STANFORD-BINET SCALE,
FORM L

by

Sister M. Fidelissima Dzik, Fel., O.S.F.

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

1952

LIFE

Sister Mary Fidelissima Dzik, Fel., O.S.F., was born in Peru, Illinois, February 2, 1900.

She was graduated from Good Counsel High School, Milwaukee, Wisconsin, June, 1918, and from the Catholic University of America, June, 1923, with the degree of Bachelor of Arts.

From 1923 to 1928 the author taught in the elementary school conducted by the Community in St. Paul, Minnesota. From 1928 to 1943 she taught Latin at Good Counsel and St. Joseph High Schools, Chicago. During the period of 1928 to 1935 she took graduate courses in Education at Loyola University, and received the degree of Master of Education, June, 1935. Since 1935 she has been engaged periodically as an instructor in Education in one of the Extension classes under the auspices of Loyola University. She undertook graduate studies in psychology at Loyola University and served an assistantship in psychology at the Loyola Center for Child Guidance, Chicago.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
Limitations of the MA and IQ scores recognized by clinical psychologists--Importance of the quality of the response on the Revised Stanford-Binet Scale--Need for an appraisal of the qualitative aspects of test responses--Value of such an ap- praisal--Bolles' evidence from sorting and group- ing tests--Thompson's proofs from tests on ability to generalize--Statement of the present problem.	
II. REVIEW OF RELATED LITERATURE	8
Relationship between the present and former studies--Studies dealing with the quantitative ana- lysis--Research works dealing with the qualitative analysis--Need for further research in the field.	
III. MATERIALS AND METHODS OF THE PRESENT STUDY	22
Subjects--Procedure for the quantitative analysis-- Development of the scales for the qualitative ana- lysis--Scoring technique--Sources of data--Some comments on the Revised Stanford-Binet Scale.	
IV. PRESENTATION AND ANALYSIS OF DATA	28
The group as a whole--Findings from the quantita- tive analysis--Findings from the qualitative analysis.	
V. SUMMARY AND CONCLUSION	60
Summary of the study--Some findings of the study.	
APPENDIX I	67
APPENDIX II	79
BIBLIOGRAPHY	85

LIST OF TABLES

Table	Page
I. DISTRIBUTION OF THE CA'S, MA'S, AND IQ'S OF THE SUBJECTS WHOSE RECORDS WERE USED IN THE PRESENT STUDY	28
II. NUMBER AND PERCENTAGE OF SUBJECTS, ACCORDING TO IQ LEVELS STUDIED, WHO PASSED THE ITEMS ON THE RE- VISED STANFORD-BINET, FORM L, YEARS V THROUGH XIV.	30
III. PERCENTAGE DISTRIBUTION OF RESPONSE RATINGS FOR ITEMS ON THE REVISED STANFORD-BINET SCALE, FORM L, YEAR V THROUGH THE AVERAGE ADULT LEVEL	39
IV. PERCENTAGE DISTRIBUTION OF RESPONSE RATINGS FOR WORDS ON THE VOCABULARY TEST OF THE REVISED STANFORD-BINET SCALE, FORM L	55

CHAPTER I

INTRODUCTION

Clinical psychologists have long recognized the fact that the mental-age and intelligence-quotient scores, in terms of which the results of mental tests are generally expressed, fail to disclose much of the valuable information about the individual's mental abilities and his probable future development that the test itself reveals. It has been their practice, therefore, to qualify the mental-age score by reference to the specific tests from which it has been derived. Moreover, they have been interested in the manner of approach, in attentiveness, in emotional attitudes, in the tendency to persevere or give up easily, and in the quality of the responses given.

The quality of the response is of great interest in a test like the Revised Stanford-Binet Scale, in which acceptable responses may range from a very simple and elementary concept to one involving a high degree of abstraction. For example, a child who defines an orange "a round reddish-yellow juicy fruit that grows in warm climates" gives evidence of far more understanding of the nature of an orange than the child who simply answers "to eat." Intelligent foreigners of limited vocabulary often

show their superiority by the quality of their definitions.

Their responses tend to be more precise and abstract than those of individuals of limited intelligence but wider vocabulary in the language used.

Quite obviously, these qualitative differences in the responses of children and, consequently, in the thought processes which give rise to them can be obscured by a system of scoring which credits all the right answers to a given item alike. Such is the case in the Revised Stanford-Binet Scale, in which each response is scored as passed or failed, depending on whether or not it satisfies the scoring standards set by the authors of the test. Thus, through a study of the qualitative aspects of children's responses, further information may be gained about the various mental abilities quantitatively evaluated.

Commenting on the value of the vocabulary test as an aid to this fuller understanding of the mental development of the child, Terman states:

This test, because it throws such interesting light on the maturity of the child's apperceptive processes, is one of the most valuable of all. It is possible to differentiate at least a half-dozen degrees of excellence in definitions, according to the intellectual maturity of the subject.¹

¹ Lewis M. Terman, The Measurement of Intelligence, Boston, 1916, 169.

Concerning the similarities or differences test, Terman says, "Its excellence lies mainly, however, in the fact that it throws light upon the character of the child's higher thought processes."² He points out that the ability to state essential similarities or differences requires a higher degree of intelligence than to give superficial, trivial, or accidental ones. He also maintains that this ability does not appear until "well toward the adult stage." On this account he considers it "a very favorable sign" if it appears in the performance of a child seven or eight years of age.³

Further objective evidence, discriminating the quality of mental functioning, comes from experimentation with tests of the sorting-and-grouping-of-objects category as contrasted with the more or less verbal tests of the Revised Stanford-Binet Scale. Bolles⁴ by means of a qualitative analysis of performance on a series of tests of this type was able to differentiate four characteristic levels of performance or "levels of abstraction" at which her groups of aments, dements, and normal children of the same MA were able to perform. These levels were: (1) identity, (2) partial identity, (3) co-functionality, and (4) categorical similarity. Grouping on the basis of identity was

2 Ibid., 201-202.

3 Ibid., 202.

4 Mary Marjorie Bolles, "The Basis of Pertinence," Archives of Psychology, XXX, No.212, June, 1937, 1-51.

designated as the most concrete and on the basis of categorical similarity as the most abstract form of behavior. She found that her groups of aments and dements always reacted in a concrete way; the normal group, consisting of children, responded in a concrete or abstract manner or both, depending on the situation. The latter group was also able to shift easily from one aspect of the situation to another, while the aments and dements found it impossible to change from a previously adopted form of behavior. Bolles maintained that these differences were not due to variations in the mental-age level because the groups were of the same mental age. However, without underestimating the value of the mental age on the Stanford-Binet, which the author regards as the best measure of "intellectual status" that we have at present, she contends, and rightly so, that from the mental age alone we cannot secure a complete and an adequate picture of the individual's mental status.

. . . It is recognized that equating for mental age on the Stanford-Binet does not necessarily equate for psychological performance on all types of tests. Taking the general level of mental function as indicated by mental age does not warrant the statement that these subjects are of the same "intellectual status."⁵

The author believes that to obtain information pertinent to the amount of abstraction attainable by a certain

5 Ibid., 45.

individual the tests must be so constituted as to permit of different levels of performance. Since most of the tests on the Revised Stanford-Binet allow for a wide range of qualitative differences in the responses, it seems reasonable to hope that much valuable information about the individual's mental status can be derived through a qualitative analysis of his responses on the test. Bolles did not seem to find the Stanford-Binet tests "as productive on the present problem as the sorting tests"; because, she claimed, the tests "depend on the highest level of abstraction. If the subject is not capable of this type of performance, the only record possible is a minus score."⁶

Thompson⁷ expresses somewhat the same view. "Items on the Stanford-Binet are scored as passed or failed, and no attempt is made to evaluate what the child is able to do if he fails the test."⁸ She drew her conclusions from a series of tests in which the ability to generalize was tested, and which were so arranged as to admit of acceptable performance at different levels. These tests were administered to school children; the younger ranging in age from six to eight, the older ranging in age from nine to

6 Ibid., 47-48.

7 Jane Thompson, "The Ability of Children of Different Grade Levels to Generalize on Sorting," Journal of Psychology, XI, January, 1941, 119-126.

8 Ibid., 124.

eleven years. Her results seemed to indicate that at all ages tested there was present in the mind of the child some basis for generalization.⁹ The younger children tended to group objects related in a specific concrete situation; whereas the older seemed to grasp basic relationships and formed categories. The latter likewise saw a wider range of possible relationships than the former.

From these considerations it is evident that the qualitative aspects of children's responses are very important not only on the sensori-motor level but in the verbal intellectual functions as well. This fact motivated the clinical psychologists to attempt analyzing qualitatively as well as quantitatively the Revised Stanford-Binet Intelligence Scale, Form L. The studies that have since been advanced afford little more than a beginning and much remains to be done.

The problem of the present study is to analyze quantitatively and qualitatively the performance of children of the same mental age but different IQ levels on the Revised Stanford-Binet Scale, Form L, in order to determine whether there is any consistency of performance among the various groups involved.

⁹ Similar results were obtained by Frances Virginia Rau in a study of conceptual thinking among young children. It is interesting to note that she found even the youngest of her subjects make some attempt at classification, thus demonstrating the use of universal concepts. Her study is entitled, The Performance of Young Children in a Classification Task, Unpublished Master's Thesis, Loyola University, Chicago, Illinois, 1943.

The study includes two related phases of the problem: (1) an analysis of the successes on the 1937 revision of the Stanford-Binet Scale; and (2) an evaluation of the contents of each reply on the basis of its relative merits according to a six-point scale, representing three degrees of a right or "plus" answer and three of a wrong or "minus" answer. The writer hopes to ascertain whether there are any differences or similarities between the groups involved in this study that are observable in a qualitative as well as quantitative analysis of performance on the items of this scale. It may even be possible to arrive at some serviceable tool for the use of the clinical psychologist who wishes a more objective standard for judging the differential quality of both the passed and failed responses on the test. The writer does not intend to suggest a new scheme of crediting responses on the test, or to find fault with the present allocation of test items at the various year levels. Such considerations are outside the scope of the present study.

CHAPTER II

REVIEW OF RELATED LITERATURE

There are numerous studies on various aspects of the Revised Stanford-Binet Scale, but comparatively few are available in which an item-by-item analysis has been made. These latter seemed especially meagre as regards the qualitative analysis of the content of test responses. Concerning similar studies, none was found to be identical insofar as the quantitative and qualitative analyses of the entire range of tests attempted by each individual are presented in the same study for a group of a restricted mental age but different IQ levels. On the contrary, it was found that some authors treated only one aspect of the present problem without differentiating the performance of the various IQ groups within the same mental-age range. Other writers varied their groups to correspond in MA with the year level considered, or confined their analysis to certain items of the test.

Since the present study is concerned with the quantitative and qualitative analysis of the test responses on the Revised Stanford-Binet Scale, research works dealing with the former type of analysis will be considered first.

Barber¹ tabulated the successes and failures for all the tests at each age level of the Revised Stanford-Binet Scale, Form L, and computed the percentage passing each test in order to investigate the relative difficulty of subtests within each year level. The results, based on data assembled from 250 test records from the files of a psychological clinic, showed that within several year levels the tests seemed to be of unequal difficulty. Such items as Picture Absurdities (VII:1), Vocabulary (VIII:1), Making Change (IX:5), Picture Absurdities (X:2), and Abstract Words (XII:5) were significantly easier for this group. On the other hand, tests of Similarities (VII:2), Similarities and Differences (VIII:4), Verbal Absurdities (IX:2), Reading and Report (X:3), Finding Reasons (X:4), and Minkus Completion (XII:6) were found to be reliably more difficult.

In a similar study of the results from 506 Stanford-Binet tests administered to children referred to a guidance agency, Gillette² secured measures of difficulty of the various test items at year levels VI through XII for three different groupings of the test data.

1 E. R. Barber, A Study of Scatter and the Relative Difficulty of Sub-tests in the Revised Stanford-Binet, Unpublished Master's Thesis, University of Illinois, Champaign, Illinois, 1938.

2 Annette L. Gillette, "Relative Difficulty of Tests within Each Year Level of the Revised Stanford-Binet, Form L, Years Six through Twelve," Journal of Psychology, XII, July, 1941, 125-138.

These groups were as follows: an MA grouping, which comprised individuals with MA's corresponding to the year level considered; a CA grouping, consisting of those testees whose CA was consistent with the year level studied; and a total group, which included all those to whom the year level was administered. She found: (1) that there were variations in difficulty between test items within each year level; (2) that the order of difficulty of test items within a year level was different for each group; and (3) that the differences between the groups were more pronounced at some levels than at others.

More pertinent, however, to the present problem is that phase of Gillette's study which compared the performance of groups of different brightness levels in an effort to ascertain the relative difficulty of test items for the low or dull, middle, and high or bright groups of the same mental age. For this purpose the mental-age groups were formed on a two-year basis; each ranging from six months below to six months above the year level considered. Thus, for example, in studying items at the six-year level of the scale, the mental-age group would extend from 5 years, 6 months to 7 years, 5 months. Each mental-age group was then divided into thirds, resulting in the groups mentioned above.

The outcomes, as measured by the percentage of children in each group passing each test, showed a variable performance on certain tests at each particular level. On the six-year level, test 4, Number Concepts, and test 3, Mutilated Pictures, were found to be easiest for the dull and middle groups, but hardest for the bright children. Test 3, Copying a Diamond, of year VII offered the least difficulty to the bright; the most, to the middle group. On the vocabulary item at year VIII, and in fact, on all the age levels, the middle group consistently surpassed both the dull and bright groups. Test 4, Similarities and Differences, showed an increase in percentages passing as the levels of intelligence were higher. The dull group found Rhymes (IX:4) the hardest; the middle group, the easiest test. Item (XI:2), Verbal Absurdities, presented little difficulty to the bright group; the most to the dull group. While no significant differences in performance were found on the twelve-year level, the middle group apparently had a slight advantage over the other two on test 3, Response to Pictures II: Messenger Boy, and Test 4, Repeating 5 Digits Reversed. Test 6, Minkus Completion, seemed equally hard to each group.

The practical value of this and similar investigations, according to the author, lies chiefly in the utilization of the findings by clinicians in a critical evaluation of a child's performance on a test.

Virginia Fleming³ had also investigated the comparative difficulty of subtests within the various age levels of the Revised Stanford-Binet Scale, Forms L and M, in a study of 210 children who took the Form L, and of 118 children who took the Form M of the revised scale. The tests were administered by graduate students, sixteen of whom gave the Form L; ten, the Form M. The results thereof were subsequently analyzed according to three methods of procedure.

An analysis of the test data, in accordance with Barber's technique, showed the tests to vary in difficulty, but very little agreement with her findings was noted.

In the second phase of her study, Fleming discussed her findings based on the procedure developed by Gillette in connection with a similar problem. In this case a fairly good agreement was found in the results at the six-year level to which this method was applied. But her findings with respect to groups of different IQ levels were not consistent with those of Gillette. Furthermore, the author observed that a slight variation in the range of the MA would yield different results.

Owing to this fluctuation in results, this investigator formulated a new technique according to which the CA's for each

3 Virginia Van Dyne Fleming, "A Study of the Subtests in the Revised Stanford-Binet Scale, Forms L and M," Journal of Genetic Psychology, LXIV, March, 1944, 3-36.

case passing and for each case failing the subtests were tabulated and the mean CA's calculated. Comparisons were then made between the mean CA of cases passing one test with the corresponding figure for another test, and between the mean CA passing and the mean CA failing the same subtest within each year level from III to IX. Her results were consistent with the findings of Terman and Merrill as regards the significance of the tests differentiating age levels. Moreover, she suggested the possibility of arranging test items in approximate order of difficulty.

Rautman⁴ reported his findings based on a study of one thousand mentally deficient individuals in connection with the problem of relative difficulty of test items. He found certain test items at a given level on the Revised Stanford-Binet definitely more difficult than others to groups of low intelligence. These included tests of verbal absurdities, sentence memories, reasoning, and picture completion of a man as compared with the vocabulary and comprehension tests which were much easier at their respective levels.

Carrying his investigation further in order to determine the influence of chronological age and experience upon test performance, this investigator subdivided each MA group on

⁴ Arthur L. Rautman, "Relative Difficulty of Test Items of the Revised Stanford-Binet; an Analysis of Records from a Low Intelligence Group," Journal of Experimental Education, X, March, 1942, 183-194.

the basis of the chronological age into younger, middle, and older groups. He found the younger children much better on tests involving pointing or some form of manual activity; such as identifying parts of the body, picture comparison, commands, bead chains, mazes, designs, paper cutting, as well as picture absurdities and the Wet Fall. On the other hand, tests of vocabulary, definitions, picture identification, and comprehension were much easier for the older children of the same mental ability. The author maintains that success on these latter tests apparently depends on age and experience. Thus he urges that the effect of CA be taken into account whenever a qualitative interpretation of a child's performance on a test is desired.

Harriman⁵ pointed out certain irregularities of performance on the Revised Stanford-Binet, Form L, for a group of two hundred fifth and sixth grade pupils, averaging 112 in IQ, and 11-7 in CA. He found test items on the thirteen-year level much easier for his subjects than those of year twelve. He also found wide discrepancies between individual items at other year levels, especially between Messenger Boy and Abstract Words of year XII; the Codes and Proverbs item on the average-adult level; the Enclosed Box Problem, Minkus Completion, and the Vocabulary item at the superior-adult one level.

⁵ Philip Lawrence Harriman, "Irregularity of Successes on the 1937 Stanford Revision of the Binet Tests," Journal of Consulting Psychology, III, May-June, 1939, 83-85.

Mitchell⁶ in her study of sixty-seven university freshmen and eighty-six senior medical students at the University of Iowa, found that year level XIII was far more difficult than either XII or XIV. Of the seventy students who passed all the items at year XIV, eighty-six per cent scored failures on the thirteen-year level; twenty-nine per cent at XII, and twenty-six per cent at both XII and XIII. Item XIII:I, Plan of Search, ranked as the most difficult test; Abstract Words as the easiest test for this group. The greatest discrepancy, however, was on the vocabulary item at the first superior-adult level, at which all the medical seniors, but only sixty-seven per cent of the university freshmen passed the test. Seventy-six per cent of the medical seniors had likewise succeeded on the vocabulary item at the third superior-adult level.

The non-verbal items at year XIII were apparently more difficult for this group than some of the verbal items at the higher levels.

Mitchell's data concur with Harriman's results inasmuch as both studies revealed irregularities at year levels XII, XIII, and XIV, and individual items at the superior-adult levels. In marked contrast to Harriman's findings, Mitchell found year XIII

⁶ Mildred B. Mitchell, "Irregularities of University Students on the Revised Stanford-Binet," Journal of Educational Psychology, XXXII, October, 1941, 513-522.

the most difficult level. Another striking difference was on the vocabulary item at the first superior-adult level, which was passed by all the medical seniors, sixty-seven per cent of the university freshmen, but failed by the children in Harriman's study. Abstract words, however, were comparatively easy for all the groups. Harriman found a very high percentage of success on the enclosed box problem and stated that it was "probably placed too high in the scale," inasmuch as it was the most frequently passed test at the superior-adult levels by his group of children. But the university students did not seem to find it very easy, for it ranked fourth in difficulty among the six tests at this level. Moreover, a comparison of the percentages of successes achieved by the two groups on some of the items of adult levels further disclosed that those items were not of equal difficulty to the two groups. For children Harriman found a marked variability as to the difficulty of those items; whereas Mitchell's results showed a fair consistency in the rank order of these same tests for her groups of adults. However, it is impossible to compare the performance of children and adults on the basis of these two studies as much as both were highly selected groups not typical of a general population.

Laycock and Clark,⁷ studying the effect of environmental factors upon test performance, analyzed the responses of matched groups of old-dull and young-bright children of a small town situated in Canada. No statistically significant differences in the type of performance of the two groups were found. But a marked superiority in favor of the old-dull was noticeable with regard to the following items: Making Change (IX:5), Picture Absurdities II (X:2), Verbal Absurdities (IX:2), and (XI:2), Comprehension IV (VIII:5), Vocabulary, Problem Situation (XI:5), and Abstract Words I (XI:3). The young-bright group consistently surpassed the old-dull on tests of immediate memory; such as memory for words, sentences, digits forward and backward, designs, and stories.

The particular kind of successes achieved by the former group was shown to depend largely on training and experience; that of the latter, on "eduction" or the individual's native ability to see meaningful relationships in the material presented. This phenomenon was particularly evidenced in the repetition of digits, for those children who depended upon "eduction", the authors observed, were able to throw them into patterns.

7 Samuel R. Laycock and Stanley Clark, "The Comparative Performance of a Group of Old-dull and Young-bright Children on Some Items of the Revised Stanford-Binet Scale of Intelligence, Form L," Journal of Educational Psychology, XXXIII, January, 1942, 1-12.

Strauss and Werner⁸ hold the distinction of being first to devise a system of evaluating the responses on the Stanford-Binet Test on the basis of the verbal logical content of the replies. According to this system ten criteria or "categories of answers" were adopted as norms for judging items involving verbal reasoning ability. At the outset of this experiment, the answers in the Pintner's Guide were first subjected to this type of scoring in order to serve later as a scoring guide. Subsequently, the responses of approximately 150 normal children, 150 high grade moron and borderline children of the endogenous type, and some 120 delinquent children of normal intelligence were matched as closely as possible with the answers in the Pintner's Guide and classified accordingly. Items of comprehension, differences, similarities, definitions, absurdities, abstract words, and fables, at year levels VI to XII inclusive on the 1916 edition of the Stanford-Binet Scale, were selected for this analysis.

The results, as measured by the percentage distributions of each type of answer, showed differences among the three groups studied in the following categories of answers: "wrong," "nonsensical," "ambiguous," and "don't know," answers. The

⁸ Alfred A. Strauss and Heinz Werner, "Qualitative Analysis of the Binet Test," American Journal of Mental Deficiency, XLV, July, 1940, 50-55.

normal children surpassed the other two groups in the number of "don't know" answers; the delinquents outnumbered the normals but not the feebleminded group in the "wrong" type of answer, while the mentally deficient group excelled in the number of "ambiguous" and "nonsensical" answers.

The authors explained the particular kind of reactions of each group by saying that the comparatively high number of "don't know" answers on the part of the normals demonstrates the power of self-criticism, a trait which is extremely weak or completely lacking among children of low intelligence. On the other hand, the intellectual defect so characteristic of the mental retardates is obvious from the high number of "nonsensical" and "ambiguous" answers.

Contrary to the above findings, Martison and Strauss,⁹ using a slightly modified form of the Strauss-Werner system of scoring test items on the 1916 edition of the Binet test, found a marked superiority of the normals over the mentally deficient children in the number of "superior" and "negative" or "don't know" answers, but not in the "superficial and nonsensical" category of answers. As to the explanation of causes, the authors considered redefinition of the norms for scoring and a somewhat inadequate sampling as the probable influencing factors.

⁹ Betty Martison and Alfred A. Strauss, "A Method of Clinical Evaluation of the Responses to the Stanford-Binet Intelligence Test," American Journal of Mental Deficiency, XLVI, July, 1941, 48-59.

Hoakley and Frazeur¹⁰ compared the performance of matched groups of endogenous and exogenous feeble-minded children on the Revised Stanford-Binet Scale to consider the possibility of securing diagnostic patterns for use in differentiating the two groups of children. The data were treated in the following manner. (1) Separate criteria for judging the diamond, the designs from memory, and the paper cutting tests were established by the authors. (2) The Terman-Merrill technique of scoring was applied to the vocabulary, abstract words, and the absurdities items. (3) In the similarity test, those items that are common to both the old and the new revision of the Binet test were scored in accordance with the Martison-Strauss standards, to which arbitrary weights were now assigned according to the order of merit indicated by those authors.

Their findings, based on the study of 18 pairs of the exogenous and endogenous type of mentally deficient children, closely matched with respect to CA, MA, and IQ, revealed no significant differences except in the case of the drawing of the diamond when scored by the Terman-Merrill method, and the memory for designs (b) test for cases within an IQ range of 65-75, regardless of the method of scoring. In view of the fact that

10 Pauline Z. Hoakley and Helen A. Frazeur, "Significance of Psychological Test Results of Exogenous and Endogenous Children," American Journal of Mental Deficiency, L, October, 1945, 263-271.

their findings were preponderantly negative and similar to Doll's results in a study of a different character, the authors caution against an indiscriminate use of diagnostic patterns in differentiating clinical cases unless thoroughly authenticated by experimentation. Even then they advise a grain of salt when considering individual cases.

The evidence pointing to consistency in performance on a test is certainly not adequate in the literature reported here. Since the data were amassed by different procedures on samplings of different populations of varying mean IQ's, the studies lack a common basis to warrant generalizations and conclusions. Furthermore, the groups in some instances were small, the findings may not be significant. Thus further investigation of the problem seems desirable.

CHAPTER III

MATERIALS AND METHODS OF THE PRESENT STUDY

The subjects of the present study were fifty-four case studies of children referred for psychological service to the Loyola Center for Guidance during the period from September, 1941, to June, 1951. The group as such is typical of a clinical population in which the proportion of children with low and high IQ's is much larger than in the general population. The reasons for their referral to the Center were many and varied; such as speech disabilities, guidance, minor and serious behavior difficulties, and school adjustment problems. The examiners whose records were studied were trained psychologists who were members of the staff or had served the Center as volunteers at various intervals since its establishment in 1941.

Of the sixty-four test records available, fifty-four were used in this study. The remaining ten had to be eliminated because some of the examiners had had to abandon the rule, generally observed at the Center, of recording oral responses verbatim. All the subjects were within the restricted range of MA 7-10 to MA 8-2 on the Revised Stanford-Binet Scale, Form L.

The group was further subdivided into three sections on the basis of the IQ. Section A included cases with IQ's of 111 and above. Section B comprised children with IQ's ranging from 90 to 110. Section C consisted of children whose IQ's were below 90. The number of subjects in each section was as follows:

Total Group¹ (MA 7-10 to MA 8-2):54

Section A (IQ 111 and over):7

Section B (IQ 90 to 110):16

Section C (IQ 89 and under):31

An individual work sheet was prepared for collecting pertinent data. It provided for copying the subject's name, sex, chronological age, mental age, intelligence quotient, examiner's name, and the date of testing. The score earned on each item and sub-item throughout the entire range of testing was also recorded, together with any comments by the examiner specifically relating to performance on individual test items.

The data were tabulated to obtain the range of chronological age and the intelligence quotient for each section of the group. The mean and median CA's, MA's, and IQ's were computed and arranged in a table to obtain a general picture of the population studied.

1 To simplify the descriptions of the various groups involved in this study the designations as indicated here will be subsequently used.

A tabulation was made showing the number of subjects in each section passing the various subtests of the Revised Stanford-Binet, Form L. The percentage of all the subjects who passed each test was calculated for each section of the group studied. The results of this quantitative analysis of the performance of each section are presented for analysis and comparison in Table II. These comparisons were drawn between the percentage of subjects in one section who passed the test and the percentage in each of the other sections who passed it.

In the second phase of this study, an attempt was made to investigate the manner or degree of superiority and inferiority with which each individual in his respective section responded to the items on the Revised Stanford-Binet Scale, Form L. To accomplish this purpose it was necessary to develop objective standards embodying the principles of scoring set up by Terman and Merrill in their manual of directions, Measuring Intelligence. A six-point scale, representing three degrees of a right or "plus" answer and three of a wrong or "minus" answer, was chosen as the means of rating the individual responses.

In constructing these scales, the following procedure seemed appropriate. In the first place, the responses of ten test records, selected at random, were transcribed and rated independently by four experienced psychologists besides the writer. Secondly, these ratings were averaged, and the resulting scores assigned to each of the responses thus evaluated.

With these criteria, twenty-one sets of scales have been developed. These scales, together with the names and location of the items on the Revised Stanford-Binet Scale, Form L, to which they were applied, are reproduced in Appendix I. An examination of the various scales makes it obvious that some items, particularly those listed in scale twenty-one, to which the responses are one-word answers, are scored for quality on a quantitative basis, i.e., the number of correct responses given.

In scoring the responses a number value of +3, +2, +1, -1, -2, or -3 was assigned to each item according to its relative quality as judged by the criteria proposed in the various scales. These scores were interpreted as follows:

Score

- +3--A superior answer, or a perfect performance
- +2--An adequate, common-sense answer, or an intermediate performance
- +1--An inferior but an acceptable answer, or a successful performance of a borderline character
- 1--A relevant but inaccurate answer, or a failure of a marginal nature
- 2--An inadequate answer, or a performance definitely below the marginal failures
- 3--A completely irrelevant or ridiculous answer, or a performance not recognizable as an attempt

This phase of the present investigation was based on an analysis of fifty-four Revised Stanford-Binet Test records, Year V through the Average Adult level. Each response was appraised on the basis of its relative merits and the rating was recorded in its appropriate space on the individual work sheet. After all the record blanks had been analyzed, tabulations were made showing the distributions of each type of answer given for each section in the group studied.

The Revised Stanford-Binet Scale is too well-known for a detailed description here. Suffice it to say that the New Revision, which has now generally replaced the single Stanford-Binet of 1916, provides two alternative scales, Forms L and M, "which differ almost completely in content, but are mutually equivalent with respect to difficulty, range, reliability, and validity."² Furthermore, the Revision eliminates many of the objectionable features of the old scale primarily by including new lower levels which are located at half-year intervals, by extending the scale upwards through the addition of two more superior adult levels, and by filling in the gaps at the eleven and thirteen years.

The number of tests has been increased from ninety in the Stanford-Binet to 129 in each of the two forms of the New

² Lewis M. Terman and Maud A. Merrill, Measuring Intelligence, Boston, 1937, 3.

Revision. The tests have also been better standardized in the new scales, since they were administered under carefully defined procedures to a group of 3,184 subjects, a sampling representative of the white population in the United States, as contrasted with the 905 subjects from the states of California and Nevada on which the standardization of the 1916 revision was made. Many tests of the old scale, which experience of two decades has shown to be unsatisfactory, were substituted by new and better ones, providing a richer sampling of mental abilities.

Following the basic Binet method, the tests are arranged according to mental-age levels--a form to which Grace H. Kent³ referred as "needlessly cumbersome and uneconomical." However, despite this and other adverse criticisms, a test of the Binet type is considered by some as "the most generally useful instrument yet devised in the field of mental measurement."⁴ Moreover, the great variety of brief tests, usually incorporated into such a scale, not only captivates the interest and enlists the cooperation of the subject but provides a greater insight into the mental development of the individual as the testing progresses from one year level to another.

3 Grace H. Kent, "Suggestions for the Next Revision of the Binet-Simon Scale," The Nineteen Forty Mental Measurements Yearbook, ed. Oscar Krisen Buros, Highland Park, New Jersey, 1941, 246.

4 Stanley D. Porteus, The Practice of Clinical Psychology, New York, 1941, 112.

CHAPTER IV

PRESENTATION AND ANALYSIS OF DATA

Table I summarizes the information concerning the composition of the groups used in the present study. For comparable purposes the total sample of fifty-four cases was subdivided into three sections on the basis of the IQ as explained in Chapter III, page 23.

TABLE I

DISTRIBUTION OF THE CA'S, MA'S, AND IQ'S OF THE SUBJECTS
WHOSE RECORDS WERE USED IN THE PRESENT STUDY

Section		Number	Range	Median	Mean
CA	A	7	6-2 to 7-3	6-11	6-8.4
	B	16	7-3 to 8-8	8- 1	8-0.8
	C	31	9-0 to 16-4	10-11	11-1.9
MA	A	7	7-10 to 8-1	7-11	7-11.1
	B	16	7-10 to 8-2	7-11	7-11.5
	C	31	7-10 to 8-2	8- 0	8-0.3
IQ	A	7	113 to 130	114	119
	B	16	91 to 109	98.5	98.8
	C	31	54 to 89	75	73.9

The table indicates that there are twice as many subjects in Section C, the slow group, as in the normal or average group, and almost five times as many as in Section A, the bright group. This somewhat greater number of subjects with low IQ's is not unusual in a clinical population. Furthermore, it is also necessary to keep in mind that the cases selected for this study were restricted to a very narrow mental-age range of seven years, ten months, to eight years, two months.

Observation of the CA scatter shows that the highest CA of the bright group, Section A, is almost two years below the lowest CA and nine years below the highest CA of the slow group, Section C. The school history records that the slow children have been in school from three to nine years, with an average period of five years; the average group, from two to four years and for an average period of 2.7 years; the young-bright group, from one and a half to two years and on an average of 1.8 years.

An examination of the IQ distributions indicates mean differences of almost 44 IQ points between the slow and bright groups and of 25 IQ points between the slow and average groups. The mean IQ's for the bright and average groups differ 19 IQ points.

Table II, pages 30-32, gives the number and percentage of successes on each test for each of the three sections. These percentages were based on the total number of subjects in each

section.¹ Hence, the lower levels of successes comprised items actually administered and passed and those for which success was presupposed because of a higher basal age achieved on the test. Similarly, the upper levels of failure included tests actually presented and failed and those for which failure was assumed by reason of two successive levels of failure.

TABLE II

NUMBER AND PERCENTAGE OF SUBJECTS, ACCORDING TO IQ LEVELS STUDIED, WHO PASSED THE ITEMS ON THE REVISED STANFORD-BINET, FORM L, YEARS V THROUGH XIV

Test Item	Description	Section A		Section B		Section C	
		No.	Per cent	No.	Per cent	No.	Per cent
V, 1	Picture Completion: Man	7	100.0	16	100.0	31	100.0
2	Paper Folding: Triangle	7	100.0	15	93.8	31	100.0
3	Definitions	7	100.0	16	100.0	31	100.0
4	Copying a Square	7	100.0	16	100.0	31	100.0
5	Memory for Sentences ^{II}	7	100.0	16	100.0	31	100.0
6	Counting Four Objects	7	100.0	16	100.0	31	100.0
VI, 1	Vocabulary	7	100.0	15	93.8	30	96.8
2	Copying a Bead Chain from Memory I	7	100.0	15	93.8	30	96.8
3	Mutilated Pictures	7	100.0	16	100.0	30	96.8
4	Number Concepts	7	100.0	16	100.0	30	96.8
5	Pictorial Likenesses and Differences	7	100.0	16	100.0	31	100.0
6	Maze Tracing	7	100.0	16	100.0	29	93.5

¹ In computing the percentages, figures to the hundredth place were used and the results rounded off to the tenth place.

TABLE II (continued)

NUMBER AND PERCENTAGE OF SUBJECTS, ACCORDING TO IQ LEVELS STUDIED,
WHO PASSED THE ITEMS ON THE REVISED STANFORD-BINET, FORM L,
YEARS V THROUGH XIV

Test Item	Description	Section A		Section B		Section C	
		No.	Per cent	No.	Per cent	No.	Per cent
VII, 1	Picture Absurdities I	4	57.1	15	93.8	27	87.1
2	Similarities:						
	Two things	6	85.7	10	62.5	21	67.7
3	Copying a Diamond	5	71.4	11	68.8	25	80.6
4	Comprehension III	5	71.4	14	87.5	27	87.1
5	Opposite Analogies I	4	57.1	16	100.0	19	61.3
6	Repeating 5 Digits	4	57.1	10	62.5	18	58.1
VIII, 1	Vocabulary	3	42.9	9	56.3	15	48.4
2	Memory for Stories:						
	The Wet Fall	6	85.7	15	93.8	22	71.0
3	Verbal Absurdities I	4	57.1	8	50.0	19	61.3
4	Similarities and Differences	5	71.4	7	43.8	9	29.0
5	Comprehension IV	5	71.4	8	50.0	21	67.7
6	Memory for Sentences III	4	57.1	8	50.0	5	16.1
IX, 1	Paper Cutting I	2	28.6	7	43.8	13	41.9
2	Verbal Absurdities II	3	42.9	4	25.0	8	25.8
3	Memory for Designs			7	43.8	11	35.5
4	Rhymes: New form	3	42.9	3	18.8	8	25.8
5	Making Change	2	28.6	3	18.8	10	32.3
6	Repeating 4 Digits Reversed	3	42.9	4	25.0	10	32.3
X, 1	Vocabulary					3	9.7
2	Picture Absurdities II: Frontier Days			7	43.8	20	64.5
3	Reading and Report					1	3.2
4	Finding Reasons I	2	28.6	3	18.8	8	25.8
5	Word Naming	1	14.3	4	25.0	12	38.7
6	Repeating 6 Digits	2	28.6	4	25.0	7	22.6

TABLE II (continued)

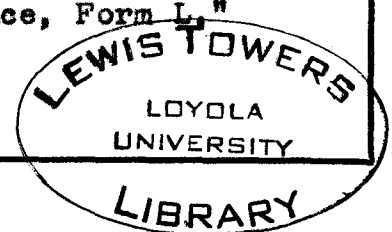
NUMBER AND PERCENTAGE OF SUBJECTS, ACCORDING TO IQ LEVELS STUDIED,
WHO PASSED THE ITEMS ON THE REVISED STANFORD-BINET, FORM L,
YEARS V THROUGH XIV

Test Item	Description	Section A		Section B		Section C	
		No.	Per cent	No.	Per cent	No.	Per cent
XI, 1 2 3 4 5 6	Memory for Designs			4	25.0	4	12.9
	Verbal Absurdities III	1	14.3				
	Abstract Words I					3	9.7
	Memory for Sentences IV	2	28.6	3	18.8	3	9.7
	Problem Situation	1	14.3			5	16.1
	Similarities: Three Things	1	14.3				
XII, 1 2 3 4 5 6	Vocabulary					1	3.2
	Verbal Absurdities II					2	6.5
	Response to Pictures II: Messenger Boy					3	9.7
	Repeating 5 Digits Reversed	1	14.3			2	6.5
	Abstract Words II					3	9.7
	Minkus Completion						
XIII, 1 2 3 4 5 6	Plan of Search					1	3.2
	Memory for Words					2	6.5
	Paper Cutting I						
	Problems of Fact	1	14.3	1	6.3	6	19.4
	Dissected Sentences						
	Copying a Bead Chain from Memory II			2	12.5	2	6.5
XIV, 1 2 3 4 5 6	Vocabulary					1	3.2
	Induction						
	Picture Absurdities III: The Shadow					1	3.2
	Ingenuity						
	Orientation: Direction I						
	Abstract Words II						

The table shows definite superiority of the slow group, Section C, over each of the other groups on test item X, 2, Picture Absurdities II: Frontier Days. Although the average group, Section B, did better by six and seven-tenths per cent at a lower level, VII, I, the slow group again was successful at the fourteen-year level, at which only one success was recorded. Laycock and Clark,² who compared the test performance of the old-dull and young-bright children, likewise found that the old-dull did better to the extent of ten per cent on item X, 2, Picture Absurdities.

On the Verbal Absurdities the slow group, Section C, was somewhat superior at the seven-year level over the other two groups; but at the nine-year level, the bright group, Section A, surpassed both groups by approximately fifteen per cent. At the higher levels, items XI, 2 and XII, 2, seemed equally difficult for each group with item XI, 2, ranking as the most difficult of the absurdities tests. Of the three sub-items which constitute this test, sub-item (a) was passed by five children, (b) was a complete failure, and (c) was a success with three children. These results differ from those of

2 Laycock and Clark, "The Comparative Performance of a Group of Old-dull and Young-bright Children on Some Items of the Revised Stanford-Binet Scale of Intelligence, Form L," Journal of Educational Psychology, XXXIII, 12.



Laycock and Clark,³ who found the old-dull to excel on the absurdities test at the higher levels.

Two other tests seemed to be decidedly easier for the slow group. One of them was item IX, 5, Making Change, resulting in approximately fifteen per cent superiority. Laycock and Clark⁴ also found a high percentage of success on this test on which only two of the forty old-dull children failed to score. Other studies of similar groups reported a close correspondence in results. The other item which appeared disproportionately high in frequency of successes at the ten-year level for the slow group was the Word Naming Test, X, 5.

Test items XI, 5, Problem Situation, and Problems of Fact, XIII, 4, were apparently easier for some of the slow children since the successes were more numerous among the slow than either the average or bright children.

Both the average and slow groups seemed to have a slight advantage over the bright group on the Comprehension Test at the seven-year level, but at the eight-year level, the bright group attained a marked superiority over the other two groups. Contradictory to the findings in the present study, Laycock and Clark⁵ reported a somewhat superior performance on the Comprehension Test at year eight on the part of the old-dull.

3 Ibid., 5.

4 Ibid., 5.

5 Ibid., 10 and 6.

The Memory for Sentences Test at year five was relatively easy for all the groups, but it differentiated sharply between the bright and slow groups at the eighth and eleventh years. A marked decrease in the performance of the slow group was noticeable at each of the two levels. Their performance likewise was comparatively inferior in the Memory for Stories: The Wet Fall. A strikingly interesting fact was observed in sub-item (a) on this test. Eighty-seven per cent of the bright group as contrasted with 18.7 per cent of the average and 35.5 per cent of the slow group succeeded in giving the correct title of the story. This outcome corroborates a previous finding by Laycock and Clark⁶ that 28 young-bright but only 8 old-dull gave the title correctly. Slight differences in the percentages passing the Memory for Digits Forward Tests were found among the three groups at all year levels. In the reversal of digits, however, the bright group was consistently superior. These results are at variance with the findings of Laycock and Clark⁷ on memory tests. They concluded that the young-bright children consistently excelled in tests of immediate memory, including the Memory for Designs, which seemed to be the most difficult test for the bright group in the present study.

6 Ibid., 11.

7 Ibid., 11.

On items, VII, 2, Similarities: Two Things, and VIII, 4, Similarities and Differences, the bright group proved markedly superior. On the latter test a distinctly higher per cent of cases passed the test as the levels of intelligence were higher. This finding is in accord with previous findings notably those of Laycock and Clark,⁸ and Gillette.⁹

The outcome on the Vocabulary Test revealed no great discrepancy in the percentage of the groups passing the test at those levels at which it occurs. Neither were the slight differences consistently in favor of the same group. When the total vocabulary scores, disregarding the year levels, were analyzed, it was found that the young-bright group averaged 7.59 words; the average group, 7.5 words; and the slow group 7.97 words.

Perhaps the poorest performance observable among the three groups under consideration was on item 3, Reading and Report, at the ten-year level. Not one of the average group scored on this test, while the slow group in spite of spending an average of five years in school achieved only one success. The test was not administered to the young-bright because of lack of sufficient school instruction.

⁸ Ibid., 12.

⁹ Gillette, "Relative Difficulty of Tests within Each Year Level of the Revised Stanford-Binet, Form L, Years Six through Twelve," Journal of Psychology, XII, 132-133.

There seems to be some uniformity in performance between the average and slow groups in item IX, 1, Paper Cutting I. The slow group, apparently found item VII, I, Copying a Diamond, easiest, the average group, the hardest. In the Memory for Designs, at both the ninth and eleventh years, however, the average group was superior. The latter also achieved a perfect performance on item VII, 5, Opposite Analogies I.

From the above analysis of the performance of the groups used, certain trends have clearly emerged, reflecting specific strengths and weaknesses which characterized these different groups in their performance on the Revised Stanford-Binet, Form L. These may be briefly summarized as follows.

1. The slow group is definitely superior in Picture Absurdities II: Frontier Days, X, 2.
2. Mental maturity seems to be the controlling factor in the vocabulary tests. Experience compensates only to a slight extent since the older-duller children did little better than the young-bright group.
3. For the most part the older-duller children have the advantage in dealing with concrete situations related to life experiences. However, when these become more subtle the young-bright children forge ahead.

4. The young-bright group is consistently superior in tests of auditory memory. This is true not only of the mere verbal memory but also in the comprehension and retention of meaningful material and the reversal of digits.

5. In tests of similarities and differences the brighter children have a decided advantage.

In the second phase of the present study, all the responses of the fifty-four subjects, whose records were used, were qualitatively analyzed by means of a six-point scale, the criteria for which were developed by the writer with the assistance of four experienced psychologists. Some of the responses to item VII, 4b, Comprehension III, together with evaluations of them are given below as an illustration of the technique used. Item VII, 4b, reads:

What's the thing for you to do when you are on your way to school and notice that you are in danger of being late?

Rating	Response	Comment
+ 3	I'd hurry up and try not to be late. Most kids run.	precise and accurate
+ 3	run as fast as you could	precise
+ 2	run	correct
+ 1	go to school. (Q) right straight.	vague, but the idea of haste may be easily inferred
- 1	go to school. (Q) that's all.	too vague and incomplete
- 2	go to the principal and tell her you're late	misunderstood
- 3	go back home; stay home	irrelevant

The results of this analysis are shown in Table III. The irregularities, noticeable in the total number of responses for the various items, arise from the fact that some of the tests in the Revised Stanford-Binet Scale are single questions while others consist of several items. Sometimes too, the child's response was not recorded in the blank; consequently it could not be evaluated.

TABLE III

PERCENTAGE DISTRIBUTION OF RESPONSE RATINGS FOR ITEMS
ON THE REVISED STANFORD-BINET SCALE, FORM L,
YEAR V THROUGH THE AVERAGE ADULT LEVEL

Test Item	Section	Number of Responses	Ratings					
			+ 3	+ 2	+ 1	- 1	- 2	- 3
			%	%	%	%	%	%
V, 1	A	2	100.0					
	B		50.0	50.0				
	C	4						
V, 2	A	1	100.0					
	B		33.3	33.3	33.3			
	C	3						
V, 3	A	6			100.0			
	B			16.7	83.3			
	C	12						
V, 4	A	6		16.7	83.3			
	B			10.0	90.0			
	C	10						

TABLE III (continued)

PERCENTAGE DISTRIBUTION OF RESPONSE RATINGS FOR ITEMS
ON THE REVISED STANFORD-BINET SCALE, FORM L,
YEAR V THROUGH THE AVERAGE ADULT LEVEL

Test Item	Section	Number of Responses	Ratings					
			+ 3	+ 2	+ 1	- 1	- 2	- 3
			%	%	%	%	%	%
V, 5	A	4		100.0				
	B	8		75.0	12.5	12.5		
	C							
V, 6	A	2	100.0					
	B	4	100.0					
	C							
VI, 2	A	7	14.3	85.7				
	B	14	21.4	71.4			7.1	
	C	28	32.1	60.7	3.6			3.6
VI, 3	A	7	71.4		28.6			
	B	14	85.7		14.3			
	C	28	78.6		17.9	3.6		
VI, 4	A	7	100.0					
	B	14	85.7		14.3			
	C	28	82.1		14.3	3.6		
VI, 5	A	7	100.0					
	B	14	100.0					
	C	28	100.0					
VI, 6	A	21	14.3	85.7				
	B	42	42.9	57.1				
	C	83	25.3	60.2	4.8		9.6	
VII, 1	A	27	29.6	25.9	11.1	3.7	14.8	14.8
	B	64	18.8	37.5	35.9		1.6	6.3
	C	123	22.0	35.0	29.3	2.4	5.7	5.7
VII, 2	A	28		17.9	42.9	14.3	21.4	3.6
	B	60		5.0	33.3	20.0	31.7	10.0
	C	121	1.7	5.0	38.8	20.7	24.0	9.9

TABLE III (continued)

PERCENTAGE DISTRIBUTION OF RESPONSE RATINGS FOR ITEMS
ON THE REVISED STANFORD-BINET SCALE, FORM L,
YEAR V THROUGH THE AVERAGE ADULT LEVEL

Test Item	Section	Number of Responses	Ratings					
			+3	+2	+1	-1	-2	-3
			%	%	%	%	%	%
VII, 3	A	21		4.8	42.9	42.9	9.5	
	B	48		4.2	58.3	31.3	6.3	
	C	93	1.1	5.4	54.8	23.7	8.6	6.5
VII, 4	A	21	19.0	28.6	19.0	4.8	28.6	
	B	48	10.4	50.0	20.8	4.2	14.6	
	C	92	9.8	50.0	8.7	3.3	27.1	1.1
VII, 5	A	7	57.1			28.6	14.3	
	B	16	93.8	6.3				
	C	31	64.5			29.0	6.5	
VII, 6	A	20		30.0	10.0	10.0	45.0	5.0
	B	43		14.0	32.6	11.6	16.3	25.6
	C	87		13.8	21.8	11.5	19.5	33.3
VIII, 2	A	7	71.4		14.3	14.3		
	B	16	12.5	75.0	6.3		6.3	
	C	31	16.1	41.9	12.9	19.4	9.7	
VIII, 3	A	27	51.9	7.4	7.4		7.4	25.9
	B	64	17.2	12.5	31.3	1.6	10.9	26.6
	C	123	23.6	13.8	23.6	6.5	8.9	23.6
VIII, 4	A	28	3.6	21.4	39.3	10.7	17.9	7.1
	B	62		8.1	40.3	37.1	6.5	8.1
	C	123		4.1	38.2	43.9	4.9	8.9
VIII, 5	A	21	9.5	33.3	14.3	4.8	23.8	14.3
	B	48	10.4	22.9	14.6	2.1	39.6	10.4
	C	91	26.4	17.6	16.5	1.1	22.0	16.5
VIII, 6	A	13		30.8	15.4		23.1	30.8
	B	24	8.3		29.2	8.3	20.8	33.3
	C	51			11.8	3.9	27.5	56.9

TABLE III (continued)

PERCENTAGE DISTRIBUTION OF RESPONSE RATINGS FOR ITEMS
ON THE REVISED STANFORD-BINET SCALE, FORM L,
YEAR V THROUGH THE AVERAGE ADULT LEVEL

Test Item	Section	Number of Responses	Ratings					
			+3	+2	+1	-1	-2	-3
			%	%	%	%	%	%
IX, 1	A	14			14.3	21.4	35.7	28.6
	B	31		3.2	19.4	22.6	35.5	19.4
	C	62		1.6	19.4	9.7	30.6	38.7
IX, 2	A	35	8.6		22.9	14.3	14.3	40.0
	B	69	2.9	11.6	23.2	2.9	21.7	37.7
	C	151	8.6	6.0	15.9	5.3	21.9	42.4
IX, 3	A	14			14.3	28.6	14.3	42.9
	B	30		16.7	20.0	6.7	36.7	20.0
	C	62		4.8	22.6	16.1	33.9	22.6
IX, 4	A	7	14.3		14.3	57.1	14.3	
	B	16			18.8	50.0	25.0	6.3
	C	31	9.7		16.1	32.3	22.6	19.4
IX, 5	A	7			28.6	14.3		57.1
	B	16	12.5		6.3	31.3		50.0
	C	31	16.1		16.1	32.3		35.5
IX, 6	A	21			14.3	38.1	19.0	28.6
	B	43			11.6	25.6	25.6	37.2
	C	87			13.8	27.6	25.3	33.3
X, 2	A	7					85.7	14.3
	B	15	20.0		26.7	6.7	33.3	13.3
	C	30	20.0	30.0	13.3		23.3	13.3
X, 3	A	11					45.5	54.5
	B	24		4.2		29.2	25.0	41.7
	C							
X, 4	A	14		21.4	14.3	64.3		
	B	29		24.1	10.3	41.4	6.9	17.2
	C	61		13.1	26.2	41.0	11.4	8.2

TABLE III (continued)

PERCENTAGE DISTRIBUTION OF RESPONSE RATINGS FOR ITEMS
ON THE REVISED STANFORD-BINET SCALE, FORM L,
YEAR V THROUGH THE AVERAGE ADULT LEVEL

Test Item	Section	Number of Responses	Ratings					
			+3	+2	+1	-1	-2	-3
X, 5	A	7	%	%	14.3	42.9	28.6	14.3
	B	16		6.3	18.8	37.5	37.5	
	C	30	3.3	6.7	30.0	20.0	33.3	6.7
X, 6	A	21			9.5	19.0	14.3	57.1
	B	43			11.6	11.6	18.6	58.1
	C	91			9.9	6.6	19.8	63.7
XI, 2	A	20	10.0	10.0		5.0	45.0	30.0
	B	40		5.0		2.5	57.5	35.0
	C	85		1.2	2.3		37.6	58.8
XI, 3	A	32			15.6	12.5	56.3	15.6
	B	60		1.7	18.3	10.0	45.0	25.0
	C	132		3.0	17.4	6.8	41.7	31.1
XI, 4	A	14		14.3	7.1	7.1	35.7	35.7
	B	25		8.0	8.0	12.0	44.0	28.0
	C	51			5.9	11.8	29.4	52.9
XI, 5	A	6	16.7				50.0	33.3
	B	15					80.0	20.0
	C	31	9.7		6.5		58.1	25.8
XI, 6	A	33	3.0	3.0	18.2	12.1	30.3	33.3
	B	57		3.5	5.3	5.3	47.4	38.6
	C	136	0.7	0.7	11.0	4.4	37.5	45.6
XII, 3	A	4				25.0	75.0	
	B	11				9.1	81.8	9.1
	C	28	3.6	7.1		21.4	60.7	7.1
XII, 4	A	12			8.3		50.0	41.7
	B	26				7.7	26.9	65.4
	C	79			2.5	3.8	17.7	75.9

TABLE III (continued)

PERCENTAGE DISTRIBUTION OF RESPONSE RATINGS FOR ITEMS
ON THE REVISED STANFORD-BINET SCALE, FORM L,
YEAR V THROUGH THE AVERAGE ADULT LEVEL

Test Item	Section	Number of Responses	Ratings					
			+3	+2	+1	-1	-2	-3
			%	%	%	%	%	%
XII, 5	A	10					70.0	30.0
	B	29			3.4	6.9	75.9	13.8
	C	91	1.1	5.5	6.6	5.5	34.1	47.3
XII, 6	A	8					37.5	62.5
	B	18			5.6		38.9	55.6
	C	90			1.1		24.4	74.4
XIII, 1	A	3					100.0	
	B	7				42.9	42.9	14.3
	C	18		5.6	5.6	33.3	38.9	16.7
XIII, 2	A	6				16.7	16.7	66.7
	B	11					27.3	72.7
	C	30			6.7	6.7	23.3	63.3
XIII, 4	A	9	22.2		11.1		22.2	44.4
	B	18	11.1	5.6	5.6	22.2	38.9	16.7
	C	47	17.0	10.6	12.8	29.8	10.6	19.1
XIII, 5	A	5					40.0	60.0
	B	4					75.0	25.0
	C	24					45.8	54.2
XIII, 6	A	2					50.0	50.0
	B	4				25.0	50.0	25.0
	C	11				36.4	18.2	45.5
XIV, 2	A	1						100.0
	B	3						100.0
	C	11						100.0
XIV, 3	A	1					100.0	
	B	3					66.7	33.3
	C	11		9.1			63.6	27.3

TABLE III (continued)

PERCENTAGE DISTRIBUTION OF RESPONSE RATINGS FOR ITEMS
ON THE REVISED STANFORD-BINET SCALE, FORM L,
YEAR V THROUGH THE AVERAGE ADULT LEVEL

Test Item	Section	Number of Responses	Ratings					
			+3	+2	+1	-1	-2	-3
			%	%	%	%	%	%
XIV, 4	A	1						100.0
	B	2					100.0	
	C	9					33.3	66.7
XIV, 5	A	1						100.0
	B	2						100.0
	C	10				20.0	20.0	60.0
A.A., 2	A							
	B							
	C	4					25.0	75.0
A.A., 3	A							
	B							
	C	15				6.7	73.3	20.0
A.A., 4	A							
	B							
	C	12						100.0
A.A., 5	A							
	B							
	C	11					54.5	45.5
A.A., 7	A							
	B							
	C	8						100.0
A.A., 8	A							
	B							
	C	21					42.9	57.1

The table indicates that on the five-year level the average group gave better responses than the slow group on four of the six tests. On item V, 3, Definitions, sixteen and seven-tenths per cent of the slow ones achieved a good response, while none of the average group rated above a marginal success. The only failures of the slowest group at this level, occurring in 12.5 per cent of the cases, were marginal failures on item V, 5, Memory for Sentences. There were no responses at this level for the young-bright since their basal ages were established at higher levels. For this reason, successes at this level were assumed.

At the six-year level the percentages of superior responses among all groups were high on all but two of the items, the Bead Chain and the Maze Tracing. The slow group, Section C, exceeded on the Bead Chain, while the average group excelled by a wide margin on the Maze Tracing test. On both of these items, however, the young-bright succeeded invariably in giving good answers unlike the other two groups who scored in inferior responses. Minus scores of a marginal character were earned by 3.6 per cent of the slow group on item VI, 3, Mutilated Pictures, and item VI, 4, Number Concepts. Nine and six-tenths per cent of the group scored below a marginal failure on the Maze Tracing Test. The average group likewise gave inadequate responses on the Bead Chain in 7.1 per cent of the cases. Three and six

tenths per cent of the slow group also scored a triple minus on the Bead Chain, characterized as a random stringing of the beads.

A closer study of the qualitative distribution at the seven-year level disclosed several differences in the performance of the three groups in each category of responses which will be discussed in the subsequent paragraphs. The young-bright were found to excel in the percentage of superior responses on the verbal items, Picture Absurdities, VII, 1, and Comprehension, VII, 4. The differences between the two remaining groups were slight. The slowest group exceeded the average group on the Picture Absurdities but not on the Comprehension test. On item VII, 5, Opposite Analogies, the average group excelled by a decidedly high per cent of superior answers.

There was little or no variation between the average and the slowest groups in the percentages of good or adequate responses, rated double plus, for item VII, 1, Picture Absurdities; item VII, 2, Similarities; item VII, 4, Comprehension; and item VII, 6, Repeating 5 Digits. But, when these two groups were compared with the young-bright group, the differences were more pronounced on items VII, 2, and VII, 6, in favor of the young-bright group and on items VII, 1, and VII, 4, in favor of the other two groups.

The percentages achieving marginal successes varied more from group to group than is the case for any other category of responses at this level. In this respect the average group rated the highest except in the test of Similarities, in which the young group exceeded.

Considerable consistency was observed among the three groups in the percentages of marginal failures, indicated by a single minus, on all items except VII, 3, Copying a Diamond, on which the young-bright group excelled by a wide margin.

A decidedly high proportion of inadequate responses, indicated by a double minus, was given by the young-bright group for item VII, 6, Repeating 5 Digits. For item VII, 1, Picture Absurdities, the young-bright group gave a comparatively high percentage of irrelevant responses.

It is apparent from an analysis of the tabulations at the eight-year level that the young-bright children were outstanding on the Memory for Stories VIII, 2, giving 71.4 per cent of superior responses and on the Verbal Absurdities, VIII, 3, with 51.9 per cent. On the Comprehension test, VIII, 5, the slow group was the most successful.

The average group, rating seventy-five per cent successful, surpassed in the good or adequate responses on the Memory for Stories test. Again the bright group scored highest.

in the percentage of good or adequate answers on item VIII, 4, Similarities and Differences; item VIII, 5, Comprehension; and item VIII, 6, Memory for Sentences.

Close agreement was noticeable in the percentages of marginal successes attained by the three groups on item VIII, 4, Similarities and Differences. Here, as well as on items VIII, 3, Verbal Absurdities, and VIII, 6, Memory for Sentences, the average group scored the highest.

Comparative evaluation of the percentages of marginal failures in the single minus column at this level revealed a striking similarity between the slow and average groups on the Similarities and Differences test. The percentages of marginal failures were high for both groups, but more so for the slow group.

All three groups achieved a high percentage of inadequate answers, indicated by a double minus, on two of the tests. The average group showed marked inferiority on the Comprehension test, VIII, 5, and the slow group on the Memory for Sentences, VIII, 6.

The percentages of irrelevant responses, scored a triple minus, on the Verbal Absurdities, VIII, 3, disclosed remarkable consistency among the groups. However, the percentages for Memory for Sentences, VIII, 6, were much higher and varied more from group to group. The lowest group gave 56.9 per cent

of irrelevant responses as contrasted with 30.8 per cent of the young-bright and 33.3 per cent of the average group.

The data at the nine-year level revealed a sharp drop in the percentages of superior responses among all three groups. Only in two items did the young-bright succeed in giving superior responses. These were: the Verbal Absurdities, IX, 2, and the Rhymes, IX, 4. On the Verbal Absurdities the percentages for the young-bright and slowest groups were identical. On the Rhymes both groups showed slightly higher percentages with the young-bright in the lead. On item IX, 5, Making Change, the slowest group surpassed the average group by a slight margin.

Good responses were given by the average and slow groups for three of the items. In these the average group was superior. The greatest difference in percentages of the marginal successes among the three groups was found on the Making Change test, IX, 5, in which 28 per cent of the young-bright as contrasted with 6.7 per cent of the average and 16.1 per cent of the slowest group achieved marginal successes.

The young-bright also tended to score higher percentages of marginal failures than either of the other two groups on four of the items. They achieved the lowest percentage of marginal failures on item IX, 5, Making Change, and were a close second to the average group on the Paper Cutting test, IX, 1.

The average and slow groups, however, exceeded by a wide margin in the percentages of inadequate responses in four of the items.

The percentages of irrelevant answers were more or less high. The young-bright gave the highest percentage of irrelevant responses on items IX, 3, Memory for Designs, and IX, 5, Making Change and the lowest in IX, 1, Paper Cutting and IX, 6, Repeating 4 Digits Reversed. The slowest group rated the highest in the percentage of irrelevant responses on the Paper Cutting test, IX, 1; the Verbal Absurdities, IX, 2; and Rhymes, IX, 4. The average group was noticeably inferior on the reversal of Digits, IX, 6.

In studying the distribution of the ratings for items at the ten-year level it was found that the slowest group had scored superior responses on the Picture Absurdities, X, 2, and the Word Naming test, X, 5. The average group likewise succeeded to give superior responses on the Verbal Absurdities, IX, 2, in 20 per cent of the cases.

All three groups succeeded in giving good responses on the Finding Reasons test, X, 4. The only good responses on the Picture Absurdities, X, 2, were achieved by the slow group in 30 per cent of the cases. Four and two-tenths per cent of the slow children also gave good responses on the Reading and Report test, X, 3. This was the only plus response attained on the test.

The slowest group had achieved marginal successes to a much higher extent than the other two groups on the Finding Reasons, X, 4, and Word Naming test, X, 5. By far the greatest percentage of marginal failures occurred on item X, 4, Finding Reasons, with the young-bright group far in the lead. On the Word Naming test the percentages of marginal failures were also high for the young-bright and average groups.

On the Picture Absurdities test X, 2, the young-bright were markedly inferior, giving 85.7 per cent of inadequate responses. This was the only test on which they failed to rise above inadequate responses. On the remaining items the distribution of inadequate responses did not differ much from that on other levels. The highest percentages of irrelevant responses appear for items, X, 3, Reading and Report, and X, 6, Repeating 6 Digits, on which more than half of each group gave irrelevant responses.

At the eleven-year level the young-bright succeeded in giving superior responses even though they had been unable to meet this standard on the preceding level. They also achieved a notable rate of good or adequate responses in three of the items. They showed the highest percentage of marginal successes for the Similarities test, XI, 6, and were not far below the other two groups in the percentages of marginal successes on the Abstract Words, XI, 3.

The percentages of inadequate responses were high for all the groups on all tests at this level. This is particularly true for the average group on item XI, 5, Problem Situation, on which 80 per cent were rated inadequate. Inadequate responses were far more frequent than irrelevant ones for both the average and young-bright groups. On item XI, 2, Verbal Absurdities; item XI, 4, Memory for Sentences; and XI, 6, Similarities, the responses of approximately fifty per cent of the slow group were appraised as irrelevant.

At the twelve-year level, only the slowest group had achieved small gains of superior and good responses in two instances. The young-bright exceeded in the percentages of marginal successes on item XII, 4, Repeating 5 Digits Reversed; and the slow ones excelled on Abstract Words, XII, 5.

The highest percentages of marginal failures were made by the young-bright and the old-dull on item XII, 3, Response to Pictures: Messenger Boy. The young-bright showed high percentages of inadequate responses at this level on all items except the Minkus Completion, on which all three groups had given mostly irrelevant responses. The average group achieved high percentage of inadequate answers on the Messenger Boy, XII, 3, and Abstract Words, XII, 5. The slow group gave the highest percentage of irrelevant responses on the reversal of digits test XII, 4, and the Minkus Completion, XII, 6.

At the thirteen-year level the distribution of the percentages of superior responses was disproportionately high for all the three groups on item XIII, 4, Problems of Fact, at which the young-bright exceeded the other two groups. The percentages of good or adequate responses as well as marginal successes were likewise high for each group on this item.

On the remaining levels, the fourteen and the average adult, the responses were chiefly inadequate or irrelevant on all items except XIV, 3, Picture Absurdities, on which a small percentage of the slow group achieved good answers. Marginal failures also occurred in a small percentage of the cases on item XIV, 5, Orientation, and A.A., 3, Differences between Abstract Words.

Table IV shows the distribution of response ratings for words on the Vocabulary test.

TABLE IV

PERCENTAGE DISTRIBUTION OF RESPONSE RATINGS FOR WORDS
ON THE VOCABULARY TEST OF THE REVISED
STANFORD-BINET SCALE, FORM L

Word	Section	No. of Responses	Ratings					
			+3	+2	+1	-1	-2	-3
			%	%	%	%	%	%
Orange	A	7		28.6	71.4			
	B	16			93.8		6.2	
	C	31		12.9	83.9		3.2	
Envelope	A	7			100.0			
	B	16			100.0			
	C	31			100.0			
Straw	A	7		14.3	85.7			
	B	16			100.0			
	C	31		3.2	93.5		3.2	
Puddle	A	7			100.0			
	B	16			100.0			
	C	31			90.3		6.5	3.2
Tap	A	7			85.7		14.3	
	B	15		6.7	93.3			
	C	30		10.0	86.7		3.3	
Gown	A	7		14.3	71.4		14.3	
	B	16			93.8		6.2	
	C	31		3.2	83.9	3.2	6.5	3.2
Eyelash	A	7			57.1	42.9		
	B	15			60.0	26.7	6.7	6.7
	C	31		3.2	54.8	41.9		
Roar	A	7		28.6	42.9	14.3	14.3	
	B	12		8.3	75.0		16.7	
	C	29		3.4	55.2		27.6	13.8
Scorch	A	6			33.3		66.7	
	B	12			16.7		50.0	33.3
	C	25			28.0	4.0	56.0	12.0

TABLE IV (continued)

PERCENTAGE DISTRIBUTION OF RESPONSE RATINGS FOR WORDS
ON THE VOCABULARY TEST OF THE REVISED
STANFORD-BINET SCALE, FORM L

Word	Section	No. of Responses	Ratings					
			+3	+2	+1	-1	-2	-3
			%	%	%	%	%	%
Muzzle	A	6					83.3	16.7
	B	11			27.3		63.6	9.1
	C	28			32.1		32.1	35.7
Haste	A	5	20.0				80.0	
	B	10					70.0	30.0
	C	28			3.6		21.4	75.0
Lecture	A	5					80.0	20.0
	B	10					80.0	20.0
	C	28		7.1	7.1		50.0	35.7
Mars	A	5				20.0	80.0	
	B	8				37.5	62.5	
	C	21		4.8	4.8	4.8	66.7	19.0
Skill	A	3					100.0	
	B	10					90.0	10.0
	C	21			4.8		33.3	61.9
Juggler	A	2			50.0		50.0	
	B	9			11.1		55.6	33.3
	C	20			10.0	5.0	60.0	25.0
Brunette	A	2					100.0	
	B	7					85.7	14.3
	C	18			27.8		50.0	22.2
Peculiarity	A	1					100.0	
	B	2					50.0	50.0
	C	7			28.6		57.1	14.3
Priceless	A	2					100.0	
	B	2					50.0	50.0
	C	9			11.1	22.2	44.4	22.2

TABLE IV (continued)

PERCENTAGE DISTRIBUTION OF RESPONSE RATINGS FOR WORDS
ON THE VOCABULARY TEST OF THE REVISED
STANFORD-BINET SCALE, FORM L

Word	Section	No. of Responses	Ratings					
			+3	+2	+1	-1	-2	-3
Regard	A		$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$
	B	5				20.0	20.0	60.0
	C	8					62.5	37.5
Disproportionate	A						50.0	50.0
	B	2				25.0	50.0	25.0
	C	4					50.0	50.0
Shrewd	A						100.0	
	B	2			20.0		60.0	20.0
	C	5						100.0
Tolerate	A	1				50.0	50.0	
	B	2					66.7	33.3
	C	3						
Stave	A						50.0	50.0
	B	2						100.0
	C	2						
Lotus	A							100.0
	B	1						100.0
	C	2						
Bewail	A							100.0
	B	1						100.0
	C	1						
Repose	A							100.0
	B	1						
	C							
Mosaic	A						100.0	
	B	1						
	C							
Flaunt	A						100.0	
	B	1						
	C							

It was observed in studying Table IV that the highest percentages achieved by all three groups on the first eight words were largely marginal successes, indicated by a single plus. A small proportion of good responses was given by the average and young-bright for some of these words, while inadequate responses were given for others. The slow group was the only one giving irrelevant responses for several of these words.

It is interesting to note the sharp drop in the percentage of marginal successes for the word "eyelash" and the sudden rise in the percentages of marginal failures when those responses which consisted in pointing to the eyebrow were evaluated. Beginning with the word "roar", however, there is a gradual drop in the percentages of marginal successes for all three groups and an increase of inadequate or irrelevant responses.

In two instances the young-bright gave irrelevant responses on this test. This was in case of the words "muzzle" and "lecture." They were also the only ones who achieved superior responses in 20 per cent of the cases on the word "haste." The slowest group continued to achieve successes of a marginal kind long after the young-bright and the average group failed to achieve any. Being older, they had the advantage of more opportunity to acquire familiarity with a greater number of words.

In the detailed analysis of the distribution of responses of each qualitative type, it was seen that no particular type of response characterized any one group exclusively. All three groups without exception scored in superior as well as inferior responses. There were variations, however, in the percentages of each type of responses given by the various groups. Each group showed marked superiority or inferiority on some of the items at all the test levels.

CHAPTER V

SUMMARY AND CONCLUSIONS

The purpose of this study was to examine the degree of consistency in performance on the Revised Stanford-Binet Scale, Form L, of groups of the same mental age but different IQ levels. By analyzing the successes and failures both quantitatively and qualitatively, the writer sought to learn whether there were any resemblances or differences among the groups involved in this study in their performance on the various items of the scale and to discover specific strengths and weaknesses characterizing these groups.

Twenty-one sets of scales were devised as a guide for the objective evaluation of the quality of both the passed and the failed responses on the test. The criteria established in these scales were determined by developing the principles of scoring set forth by Terman and Merrill in their manual of directions, Measuring Intelligence. Accordingly, a scale for the qualitative evaluation of responses was constructed with the aid of four experienced clinical psychologists.

Fifty-four records of the Revised Stanford-Binet Scale, Form L, from the files of the Loyola Center for Guidance

were available for this study. These were grouped on the basis of the IQ into three sections and subjected to a double analysis. In the quantitative analysis, the percentage of all the successes on each test was determined for each section separately. The qualitative analysis consisted of rating each individual response according to a six-point scale of values, representing three degrees of a right answer and three of a wrong answer. The criteria for this analysis were proposed in the various scales developed for this purpose. For each type of analysis, the data of the findings were tabulated separately and analyzed by comparing the percentages achieved by individual groups on the various tests of the Revised Stanford-Binet Scale.

Previous studies dealt chiefly with the quantitative analysis of the individual successes on the test for groups whose mental ages coincided with the age level considered but did not differentiate the performance of the various IQ groups within each mental-age range. Likewise, research treatises based on the qualitative analysis of the child's responses on the test were confined to similar mental-age groups and limited to certain types of items. In the present study, however, the performance of the total group of 54 cases was analyzed quantitatively and qualitatively on the entire range of tests attempted by each individual in his respective IQ group.

The results of the quantitative analysis indicated that there is some, though slight, consistency in the performance of the three IQ groups ranging in mental age from 7-10 to 8-2. Most outstanding consistency observed was on item 2, Memory for Stories: The Wet Fall, at the eight-year level. It was apparently the easiest item for all three groups since it was passed by the highest percentage of cases in each group at that level. This consistency also appeared in tests: VIII, 4, Similarities and Differences; VIII, 6, Memory for Sentences III; XI, 4, Memory for Sentences IV; and XI, 6, Similarities: Three Things. The percentage of successes increased with the higher intelligence quotient--thus corroborating the findings of previous studies.

On the Picture Absurdities, item X, 2, the slow group, Section C, surpassed the other two groups by a wide margin. The average group, however, exceeded both groups at a lower level, VII, 1.

In general, on items dealing with practical life situations, namely, Comprehension III and IV, and the Verbal Absurdities I, II, and III, the slow had an advantage at the lower levels, but in proportion to the increased difficulty of the items the young-bright group forged ahead.

Some items which have been shown to depend on age and experience were consistently easier to the slow group according to the percentages passing. Belonging in this category were the following items: IX, 5, Making Change; X, 5, Word Naming; XI, 5, Problem Situation; and XIII, 4, Problems of Fact. Other studies likewise have reported on the items a high frequency of successes among the slow children.

A close similarity was noticeable in the percentages of the groups passing the vocabulary test. On closer scrutiny, it became evident that mental maturity seemed to be the controlling factor on this item. Experience compensated only to a slight extent since the slow group did somewhat better than the young-bright group.

The average and the slow groups had a slight advantage over the young-bright on items dealing with hand-eye coordinations at those age levels in which they occurred.

The findings based on the qualitative analysis of the content of each reply have shown that, whereas no particular type of response characterized any one group exclusively, variations in the percentages of each type of response reflect qualitative differences in the performance.

In analyzing the results at the various year levels, particularly noteworthy were the findings at the eight-year level. Of the six items at this level, item 2, Memory for

stories: The Wet Fall, was passed by the highest percentage of cases in each group--the young-bright surpassing the other two. The qualitative analysis on the same item showed the young-bright leading by a wide margin in the superior responses; and the average group, in the adequate, common-sense responses. The young-bright demonstrated their superior proficiency over the other two groups also on item 3, Verbal Absurdities, having scored 51.9 per cent superior responses. Item 5, Comprehension IV, showed that the highest percentages of good or adequate responses were gained by the young-bright ones. On item 6, Memory for Sentences, the average group alone scored in superior responses; and the young-bright, in good responses. The slow group not only failed to rise above the marginal successes on this item but also scored highest in irrelevant responses.

It was further observed that the slow group persisted in giving irrelevant responses in the highest number of cases on some of the memory items at other age levels. Such was the case for item XI, 4, Memory for Sentences. On the digits series, both forward and reversed, similar results in the percentages of irrelevant responses were found for the slow group with but one exception. As might be expected, an irrelevant response on the above items would be a series quite unlike the original one.

such a response in the words of Terman "is an unfavorable sign, indicating weakness of the critical sense which is so often found with low-level intelligence."¹

The results of the quantitative analysis revealed a disproportionately high frequency of successes among the slow children on item X, 2, Picture Absurdities II: Frontier Days, confirming the findings in previous studies. The qualitative analysis showed similarly a high proportion of superior and good responses on the part of the slow group after a total failure or only slight gains on the preceding lower age levels.

On the whole, the performance on the vocabulary items seemed to be characterized by marked mediocrity with a slight tendency for the young-bright children to give a definition of better quality more often than the other two groups.

CONCLUSIONS

1. It seems valuable for a critical evaluation of test performance to consider the qualitative as well as the quantitative aspects of the child's responses.
2. In this study, the quantitative analysis disclosed characteristic strengths and weaknesses of the various groups on the test. The young-bright were generally superior in items dependent on verbal reasoning ability and items of immediate

1 Terman, The Measurement of Intelligence, 195.

memory, both rote and logical; the slow, in tests dealing with simple concrete situations, and items dependent upon extended life experience.

3. Mental maturity seemed to be the deciding factor for successes on the vocabulary test. Experience accounted for the slightly higher gains on the part of the old-dull.
4. Marked discrepancies in difficulty were noticeable between test levels and between test items on a single level.
5. The qualitative analysis showed that the young-bright tend to excel in the quality of their responses on tests of verbal reasoning ability.
6. The slow group seemed to give markedly inferior responses in tests of immediate memory, chiefly memory for sentences and the digits.
7. Possibilities for the use of the various scales as an objective guide in appraising the differential quality of test responses in a clinical setup can be recognized. In this capacity, they can help the clinical psychologist decide whether the test in question has been an adequate measure of the mental development of the person concerned.
8. Further investigation of the problem at other mental-age ranges and with larger groups seems desirable.

APPENDIX I

CRITERIA FOR EVALUATING THE RELATIVE QUALITY OF RESPONSES

1. Suggested Norms for the "Verbal Items."

A response based on a clear understanding of the problem and stated in a precise, accurate, and comprehensive way was scored as triple plus.

A response not as good as the above, but a generally correct response which might be characterized by an apparently imperfect grasp of the problem, indefiniteness, or lack of some comprehensiveness was scored as a double plus.

An answer that was either incomplete or vague, but accepted by scoring standards was scored single plus. In this category were included answers "by gesture or example," and those elicited by further questioning, provided these were acceptable.

Answers that were partly correct, but were too incomplete or inaccurate to be given any credit, were rated single minus.

Answers that were wrong in the sense of being contradictory to stated facts; those based on a misunderstanding of the question or confused with something else; refused and "I don't know" answers, were given a rating of a double minus.

Answers that were completely irrelevant because they introduced factors that were beside the point and those whose content was absurd or meaningless were scored triple minus.

Name and Location of Items Scored by Above Criteria

Definitions - V, 3.

Vocabulary - 5 year levels

Picture absurdities - VII, 1; X, 2; XIV, 3.

Similarities - VII, 2; XI, 6.

Comprehension - VII, 4; VIII, 5.

Verbal absurdities - VIII, 3; IX, 2; XI, 2.

APPENDIX I (continued)

Similarities and Differences - VIII, 4.
Finding Reasons - X, 4.
Abstract words - XI, 3; XII, 5; XIV, 6.
Response to pictures - XII, 3.
Problem situation - XI, 5.
Problems of fact - XIII, 4.
Proverbs - A.A., 5.
Reconciliation of opposites - A.A., 8.

2. Suggested Norms for "Memory for Sentences."

No triple plus was given unless the psychologist commented upon superior performance and both sentences were perfectly repeated.

A double plus was given for each sentence if both were perfectly repeated.

A correct response short of the above was rated as single plus.

A rating of a minus one was given for transposition or substitution of words without change of meaning or grammatical variation.

A double minus was given for omission of words or phrases.

A triple minus was given for omission of material to the extent that meaning was destroyed or changed, or for insertion of irrelevant material, or for refusals.

Name and Location of Items Scored by Above Criteria

Memory for Sentences - V, 5; VIII, 6; XI, 4; A.A., 7.

3. Suggested Norms for "Memory for Digits and for Words"

No triple plus was given unless the psychologist commented on superiority of performance and all of the two or three items at a level were given correctly.

Each item was scored a double plus if all of the two or three items at a level were given correctly.

APPENDIX I (continued)

A correct response short of the above was scored as single plus.

For a single transposition within a series, a rating of single minus was given.

A double minus was given for more than one transposition, or for any single omission or substitution.

For more than one omission or substitution, or for refusals, a triple minus was given.

Name and Location of Items Scored by Above Criteria

Repeating 5 digits - VII, 6.

Repeating 4 digits reversed - IX, 6.

Repeating 6 digits - X, 6.

Repeating 5 digits reversed - XII, 4.

Memory for Words - XIII, 2.

4. Suggested Norms for the "Bead Chain"

A triple plus was given for a rapid performance (one minute or less) or one that was accompanied by a favorable comment from the psychologist.

A successful performance short of the above was scored as double plus.

For a spontaneous correction, or performances completed within the full allotment of time (2'), a score of single plus was given.

Failure to complete the pattern within the allotted time (2'), or a single omission or substitution provided more than half of the pattern was reproduced correctly was scored a single minus.

Refusals or reproductions definitely below marginal failures but with some features of the original pattern still observable were scored as double minus.

For any random stringing of beads a score of triple minus was given.

APPENDIX I (continued)

Name and Location of Tests Scored by Above Criteria

Copying a bead chain from memory I-VI, 2; II-XIII, 6.

5. Suggested Norms for "Memory for Stories" (VIII, 2)

A score of triple plus was given if credit was earned on each of the items of the test.

If credit was earned for each item on the test except for the title, item (a), a double plus was given.

The item was scored single plus if the performance satisfied the minimum requirements for passing, 5 plus responses.

If three or four responses were correct, the item was scored single minus.

A double minus was given for a performance in which less than three correct responses were given.

Complete failure was scored triple minus.

6. Suggested Norms for "Reading and Report" (X, 3)

If recall was perfect (21 memories) or nearly so (18 memories) and the selection read without a mistake, a triple plus was given.

An intermediate successful performance somewhat short of the above in those respects on which success depends was scored a double plus.

If performance satisfied the minimum requirements for passing the test--Memories 10, Time 35", Mistakes 2--it was scored single plus.

Single minus was given for 10 or more memories, more than 35" for reading, or more than 2 errors.

APPENDIX I (continued)

Double minus was given for less than ten memories.

Triple minus was given for less than five memories.

7. Suggested Norms for "Paper folding: Triangle" (V, 2)

No triple plus was given unless the psychologist commented on superiority of performance and the resulting figure was a perfect triangle.

An intermediate successful performance was scored double plus.

A successful performance which approximated to sample as given in Terman-Merrill manual, p. 218, was rated single plus.

Marginal failures (performances in which the correct procedure was used, but the resulting figure was not triangular in shape) were rated as single minus.

A double minus was given for refusals and for incorrect responses provided there was indication of some success.

A triple minus was given for any haphazard or random folding of the paper.

8. Suggested Norms for the "Maze tracing" test (VI, 6)

A score of triple plus was given if the marking was perfect or nearly so.

Intermediate successful performances were scored double plus.

Marginal successes that met the plus requirements as described in the Terman-Merrill manual (p. 225) were scored single plus.

Marginal failures were scored single minus.

Refusals or choice of the incorrect path was scored double minus.

Any random marking of the paper was scored triple minus.

APPENDIX I (continued)

9. Norms for the "Picture Completion: Man" Test (V, 1)

Drawings receiving three credits according to Terman-Merrill were given a triple plus.

A double plus was given for drawings receiving two credits according to Terman-Merrill.

Drawings receiving one credit according to Terman-Merrill manual were rated single plus.

Drawings omitting essential features but showing nonessential ones, such as ears, hat, hair, eyelashes, eyebrows, etc., were scored single minus.

A double minus was given for refusals.

A triple minus was given for performances approximating to samples of no-credit drawings as given in Terman-Merrill manual, p. 211.

10. Suggested Norms for Scoring Diamonds and Squares

If performance approximated to samples of marginal successes or failures as given in Terman-Merrill manual, it was scored as single plus or single minus respectively.

If drawn perfectly or nearly so, it was scored a triple plus.

Intermediate successful performances were given a double plus.

Random or highly erratic drawings were scored as triple minus, and those definitely below the marginal failures but with some semblance of four sides and four angles: a double minus.

Name and Location of Tests

Copying a square - V, 4.

Copying a diamond - VII, 3.

APPENDIX I (continued)

11. Suggested Norms for "Paper Cutting" Test (IX, 1; XIII, 3)

If the drawing was perfect or nearly so, it was scored a triple plus.

Intermediate successful performances were scored as double plus.

For drawings which approximated to samples scored plus according to Terman-Merrill manual, a rating of single plus was given.

Those drawings in which the cut was a square (triangle, circle), or in which the cut was not on the crease or not intersected by the crease (s) were scored single minus.

Failure to indicate both crease(s) and cut correctly, or omission of either cut or crease(s) was scored double minus.

Random or highly erratic drawings were scored triple minus.

12. Suggested Norms for "Designs from Memory" Test (IX, 3; XI, 1)

If the design was drawn perfectly or nearly so, the item was scored triple plus.

Intermediate successful performances receiving full credit according to Terman-Merrill were scored double plus.

Marginal successes for which full credit was earned were scored single plus; or if two half-credits were earned, each was rated a single plus.

Performances for which one half-credit was earned were rated single minus.

If the performance approximated to samples of failure as given in Terman-Merrill manual, it was scored double minus.

Random or highly erratic drawings were scored triple minus.

APPENDIX I (continued)

13. Suggested Norms for "Plan of Search" (XIII, 1)

A triple plus was given for performances in which the execution of the plan was perfect or nearly so, and if the chosen plan conformed to one of the superior types discussed in Terman-Merrill manual, samples 1, 2, 3, 4, 9, and 10.

An intermediate successful performance was scored double plus. (Sample 5 in Terman-Merrill manual).

A single plus was given for performances showing that the underlying principle was grasped, but its execution was poor, or if performance approximated to sample 6, 7, and 8 in Terman-Merrill manual.

A rating of single minus was given for performances which approximated to samples of failure as given in Terman-Merrill manual.

A performance characterized by an absence of a definite plan was scored double minus.

Refusals or highly erratic drawings were scored triple minus.

14. Suggested Norms for the "Induction" Test (XIV, 2)

Performances in which the correct rule was given spontaneously before the completion of the test were scored a triple plus.

A double plus was given if all of the six items were passed and the rule was stated satisfactorily.

A successful performance which satisfied the minimum standards for scoring was rated as a single plus.

A score of a single minus was given if subject scored on item (f) of the test, but, in stating the rule, merely recapitulated the correct number of holes for each folding, or gave an explanation which, though inadequate, did not contradict the underlying principle.

A double minus was given for refusals, or for incorrect statement of the rule; or failure on item (f).

APPENDIX I (continued)

A score of a triple minus was given for performances characterized by a complete failure to grasp the principle, as evidenced by incorrect responses on the last three items, or for any meaningless or absurd statement of the rule.

15. Suggested Norms for the "Arithmetical Reasoning" Test (A.A., 4)

A triple plus was given for each if all of the three problems were solved correctly.

Each item was scored double plus if two problems were solved correctly and the third solved after rereading.

Correct responses short of the above were scored single plus.

If only one answer was correct and two were incorrect, the incorrect responses were scored single minus.

If two problems were solved incorrectly and \$1.05 was given as answer for the third and for refusals, a double minus was given.

A triple minus was given for each if all of the three problems were solved incorrectly.

16. Suggested Norms for Ingenuity item (XIV, 4; A.A., 6)

A triple plus was given for each if all of the three problems were solved correctly.

A double plus was given for each if two problems were solved correctly.

A single plus was given if one problem was solved correctly.

If no credit was earned, but the problem was solved with extension of time, the response was marked a single minus.

Refusals or groping responses were scored double minus.

Random or irrelevant responses were scored triple minus.

APPENDIX I (continued)

17. Suggested Norms for the "Word Naming" Test (X, 5)

If 45 or more words were given, or if the list consisted of unusual or abstract words, or was characterized by richness of associations, a score of triple plus was given.

A double plus was given for a list of 35 to 44 words, or if the list consisted of "objects immediately present" and words suggested by them.

Successful performances consisting of a list of 28 to 34 isolated words were scored single plus.

If the response just barely missed being acceptable (20 to 27 words), it was scored a single minus.

If less than twenty words were named, a double minus was given.

If less than half (14) words were given, or if the list was characterized by numerous repetitions, and for refusals, a triple minus was given.

18. Suggested Norms for the "Minkus Completion" Test (XII, 6)

A triple plus was given for each if the psychologist commented on rapid or superior performance and credit was earned on each of the four trials.

Each correct item was scored double plus if at least three of the four sentences were correctly completed.

Successful performances short of the above were scored single plus.

Refusals or partial successes on the fourth sentence were rated single minus.

Sentences in which the connecting word failed to make perfect sense were scored double minus.

Random or irrelevant responses were scored triple plus.

APPENDIX I (continued)

19. Suggested Norms for "Dissected Sentences" (XIII, 5)

A superior performance indicated as such by the psychologist, or if all of the three sentences were given correctly, was scored triple plus.

Each trial for which full credit was granted was scored double plus.

If two half-credits were earned, each was rated a single plus; or if a full credit was earned after one or more spontaneous corrections were made.

A single minus was given for performances extended beyond the given limits of time; or a single half-credit earned.

A double minus was given for responses in which the order of words did not make perfect sense; for omission or substitution of more than one word, for complete alteration of the meaning intended; and for refusals.

A score of a triple minus was given for any random, non-sensical rearrangement of words.

20. Suggested Scoring for "Codes" Item (A.A., 2)

If full credit was earned for each of the two codes, each was scored as triple plus.

If full credit was earned for either of the codes, that part received double plus.

If two half-credits were earned, each was credited a single plus.

If no credit was earned, but comprehension of system was present but with confusion of alphabet (evidenced by having half or more correct, or by psychologist's notation) a score of single minus was given.

Refusals or groping responses were scored double minus.

APPENDIX I (continued)

Irrelevant or random responses were rated as triple minus.

21. Norms for the following list of items:

Name and Location of Items:

Counting four objects - V, 6.
Number concepts - VI, 4.
Making change - IX, 5.
Mutilated pictures - VI, 3.
Pictorial likenesses and differences - VI, 5.
Opposite analogies I - VII, 5.
Rhymes: new form - IX, 4.
Orientation: Directions I - XIV, 5.

If all the items of a given test were answered correctly, the test was scored a triple plus.

If the performance on a given test exceeded the minimum requirements for passing, it was scored a double plus.

A successful performance which satisfied the minimum requirements for passing was scored a single plus.

A single minus was given for performances that approximated the scoring standards but fell slightly below requirements.

Refusals were scored double minus.

If none of the responses on a given test were correct, the item was scored triple minus.

APPENDIX II

DISTRIBUTION OF RESPONSE RATINGS FOR ITEMS RANGING FROM THE FIFTH YEAR THROUGH THE AVERAGE-ADULT LEVEL OF THE REVISED STANFORD-BINET SCALE, FORM L

Test Item	T N	Section A						Section B						Section C					
		+3	+2	+1	-1	-2	-3	+3	+2	+1	-1	-2	-3	+3	+2	+1	-1	-2	-3
V, 1	7							2						2	2				
V, 2	5							1						1	1	1			
V, 3a	6									2					1	3			
V, 3b	6									2						4			
V, 3c	6									2					1	3			
V, 4a	6									2						4			
V, 4b	5								1	1					1	2			
V, 4c	5									2						3			
V, 5a	6								2						3	1			
V, 5b	6								2						3		1		
V, 6	6							2						4					
VI, 2	49	1	6					3	10			1		9	17	1			1
VI, 3	49	5		2				12		2				22		5	1		
VI, 4	49	7						12		2				23		4	1		
VI, 5	49	7						14						28					
VI, 6a	49	1	6					6	8					7	18	1		2	
VI, 6b	49	1	6					6	8					7	15	1		5	
VI, 6c	48	1	6					6	8					7	17	2		1	
VII, 1a	53	1	3	1			1	2	12	2				4	18	8	1		
VII, 1b	54	1	2			3	1	1	2	11			2	6	10	11		2	2
VII, 1c	54	2	1	1	1	1	1	3	3	8		1	1	2	8	14	1	4	2
VII, 1d	53	4	1	1			1	6	7	3				15	7	3	1	1	3

APPENDIX II (continued)

DISTRIBUTION OF RESPONSE RATINGS FOR ITEMS RANGING FROM THE
FIFTH YEAR THROUGH THE AVERAGE-ADULT LEVEL OF THE
REVISED STANFORD-BINET SCALE, FORM L

Test Item	T N	Section A						Section B						Section C					
		+3	+2	+1	-1	-2	-3	+3	+2	+1	-1	-2	-3	+3	+2	+1	-1	-2	-3
VII, 2a	52		2	2	1	2			1	5	4	4		1	1	18	1	8	2
VII, 2b	54			7						9	2	5		1	2	20	5	3	
VII, 2c	54		3	3	1				2	6	2	3	3		3	8	15	5	
VII, 2d	49				2	4	1				4	7	3			1	4	13	10
VII, 3a	54			4	1	2				10	5	1			3	18	6	2	2
VII, 3b	54		1	1	5				1	10	4	1		1	2	12	11	3	2
VII, 3c	54			4	3				1	8	6	1				21	5	3	2
VII, 4a	54	3	2	1	1			3	4	4	2	3		3	19	5	2	2	
VII, 4b	54		3	1		3		1	11	2		2		2	13		1	14	1
VII, 4c	53	1	1	2		3		1	9	4		2		4	14	3		9	
VII, 5	54	4			2	1		15	1					20			9	2	
VII, 6a	52		2	1	1	3			2	5	3	2	2		4	8	5	5	9
VII, 6b	49		2	1	1	2			2	4	1	4	4		4	7	1	8	8
VII, 6c	49		2			4	1		2	5	1	1	5		4	4	4	4	12
VIII, 2	54	5		1	1			2	12	1		1		5	13	4	6	3	
VIII, 3a	53	3					3		2	7		2	5	6	5	10	4	4	2
VIII, 3b	54	4	1	1			1	3	2	7		1	3	10	1	13	1	1	5
VIII, 3c	54	2	1	1			3		3	4	1	1	7	4	5	4	2	5	11
VIII, 3d	53	5				2		8	1	2		3	2	9	6	2	1	1	11
VIII, 4a	54		2	2	1	1	1		3	7	5	1			3	17	10	1	
VIII, 4b	54	1		5			1			10	5		1		1	18	10	1	1
VIII, 4c	53		3		4				2	2	6	2	3		1	5	15	3	7
VIII, 4d	52		1	4	2					6	7	1	1			7	19	1	3
VIII, 5a	54	1		1	1	2	2	4	1			9	2	11		4		5	11
VIII, 5b	54	1	2			3	1		3	2		9	2	5	6	1	15		4
VIII, 5c	52		5	2				1	7	5	1	1	1	8	10	10	1		

APPENDIX II (continued)

DISTRIBUTION OF RESPONSE RATINGS FOR ITEMS RANGING FROM THE
FIFTH YEAR THROUGH THE AVERAGE-ADULT LEVEL OF THE
REVISED STANFORD-BINET SCALE, FORM L

Test Item	T N	Section A						Section B						Section C					
		+3	+2	+1	-1	-2	-3	+3	+2	+1	-1	-2	-3	+3	+2	+1	-1	-2	-3
VIII, 6a	44		2	2			2	1		6	2	1	4			3	2	3	16
VIII, 6b	44		2			3	2	1		1		4	4			3		11	13
IX, 1a	54			2	2	1	2		1	6	5	2	2		1	12	6	6	6
IX, 1b	53				1	4	2				2	9	4				13	18	
IX, 2a	54	1		4	1		1		3	6		2	5	2	2	9	3	5	10
IX, 2b	52	1		2		2	2	1	2	5		1	5	7	1	6	1	5	11
IX, 2c	53			1		1	5	1	3	4	2	1	4	1	5	6		5	14
IX, 2d	47			1	2	1	3			1		4	6	3	1	3	10	12	
IX, 2e	49	1			2	1	3					7	6				4	8	17
IX, 3a	53			1	3	2	1		3	2	2	6	2		1	6	5	14	5
IX, 3b	53			1	1		5		2	4		5	4		2	8	5	7	9
IX, 4	54	1		1	4	1				3	8	4	1	3		5	10	7	6
IX, 5	54			2	1		4	2		1	5		8	5		5	10		11
IX, 6a	49			1	4		2				4	3	5			5	8	5	12
IX, 6b	52			1	2	1	3			2	3	4	7			4	7	9	9
IX, 6c	50			1	2	3	1			3	4	4	4			3	9	8	8
X, 2	52					6	1	3		4	1	5	2	6	9	4		7	4
X, 3	35											5	6		1		7	6	10
X, 4a	53		2	1	4				5	2	5		3		2	10	14	5	
X, 4b	51		1	1	5				2	1	7	2	2		6	6	11	2	5
X, 5	53			1	3	2	1		1	3	6	6		1	2	9	6	10	2

APPENDIX II (continued)

DISTRIBUTION OF RESPONSE RATINGS FOR ITEMS RANGING FROM THE
FIFTH YEAR THROUGH THE AVERAGE-ADULT LEVEL OF THE
REVISED STANFORD-BINET SCALE, FORM L

Test Item	T N	Section A						Section B						Section C					
		+3	+2	+1	-1	-2	-3	+3	+2	+1	-1	-2	-3	+3	+2	+1	-1	-2	-3
X, 6a	53			2		2	3			3	2	4	6			2	2	2	25
X, 6b	51				1	1	5			1	1	2	10			3	2	5	20
X, 6c	51				3		4			1	2	2	9			4	2	11	13
XI, 2a	51	1	1		1	2	2				1	8	5		1	2		9	18
XI, 2b	51					4	3		1			10	4					15	14
XI, 2c	43	1	1			3	1		1			5	5					8	18
XI, 3a	48			3	1	3				5	1	4	2		2	13	5	6	3
XI, 3b	47					4	2			2		8	1			2		16	12
XI, 3c	46					5	2					6	6			3		12	12
XI, 3d	49			2	3	2			1	4	5	5	6		1	5	4	14	3
XI, 3e	34					4	1					4	6		1			7	11
XI, 4a	44		1	1		4	1		1	2	1	8				3	5	11	6
XI, 4b	46		1		1	1	4		1		2	3	7				1	4	21
XI, 5	52	1				3	2					12	3	3		2		18	8
XI, 6a	47			3		2	2			1		7	2			6		12	12
XI, 6b	51		1	1		2	3		2			6	5		1	4		10	16
XI, 6c	51			2	1	2	2			2		6	6			4	1	11	14
XI, 6d	41	1					4				2	4	5				3	9	13
XI, 6e	36				3	4					1	4	4	1		1	2	9	7
XII, 3	43				1	3					1	9	1	1	2		6	17	2
XII, 4a	40			1		2	1				2	1	6			2	1	6	18
XII, 4b	39					2	2					1	7				1	5	21
XII, 4c	38					2	2					5	4				1	3	21
XII, 5a	30					2						7	1				1	9	10
XII, 5b	35					2	1					6	1		4			9	12
XII, 5c	35					3					2	5	2			3	4	5	10
XII, 5d	30						2		1			4		1		3		8	11

APPENDIX II (continued)

DISTRIBUTION OF RESPONSE RATINGS FOR ITEMS RANGING FROM THE
FIFTH YEAR THROUGH THE AVERAGE-ADULT LEVEL OF THE
REVISED STANFORD-BINET SCALE, FORM L

[illegible]

BIBLIOGRAPHY

I. BOOKS

- McNemar, Quinn, The Revision of the Stanford-Binet Scale; An Analysis of the Standardization Data, Boston, 1942.
- Pintner, Rudolph, Anna Dragositz, and Rose Kushner, Supplementary Guide for the Revised Stanford-Binet Scale (Form L), Applied Psychology Monographs of the American Psychological Association, No. 3, Stanford University, California, 1944.
- Porteus, Stanley D., The Practice of Clinical Psychology, New York, 1941.
- Terman, Lewis M., The Measurement of Intelligence, Boston, 1916.
- Terman, Lewis M., and Maud A. Merrill, Measuring Intelligence, Boston, 1937.

II. ARTICLES

- Bolles, Mary Marjorie, "The Basis of Pertinence," Archives of Psychology, XXX, No. 212, June, 1937, 1-51.
- Fleming, Virginia Van Dyne, "A Study of the Subtests in the Revised Stanford-Binet Scale, Forms Land M," Journal of Genetic Psychology, LXIV, March, 1944, 3-36.
- Gillette, Annette L., "Relative Difficulty of Tests within Each Year Level of the Revised Stanford-Binet, Form L, Years Six through Twelve," Journal of Psychology, XII, July, 1941, 125-138.
- Harriman, Philip Lawrence, "Irregularity of Successes on the 1937 Stanford Revision of the Binet Tests," Journal of Consulting Psychology, III, May-June, 1939, 83-85.

- Hoakley, Pauline A., and Helen A. Frazeur, "Significance of Psychological Test Results of Exogenous and Endogenous Children," American Journal of Mental Deficiency, L, October, 1945, 263-271.
- Kent, Grace H., "Suggestions for the Next Revision of the Binet-Simon Scale," The Nineteen Forty Mental Measurements Yearbook, edited by Oscar Krisen Buros, Highland Park, New Jersey, 1941, 246-247.
- Laycock, Samuel R., and Stanley Clark, "The Comparative Performance of a Group of Old-dull and Young-bright Children on Some Items of the Revised Stanford-Binet Scale of Intelligence, Form L," Journal of Educational Psychology, XXXIII, January, 1942, 1-12.
- Martison, Betty, and Alfred A. Strauss, "A Method of Clinical Evaluation of the Responses to the Stanford-Binet Intelligence Test," American Journal of Mental Deficiency, XLVI, July, 1941, 48-59.
- Mitchell, Mildred B., "Irregularities of University Students on the Revised Stanford-Binet," Journal of Educational Psychology, XXXII, October, 1941, 513-522.
- Rautman, Arthur L., "Relative Difficulty of Test Items of the Revised Stanford-Binet; An Analysis of Records from a Low Intelligence Group," Journal of Experimental Education, X, March, 1942, 183-194.
- Strauss, Alfred A., and Heinz Werner, "Qualitative Analysis of the Binet Test," American Journal of Mental Deficiency, XLV, July, 1940, 50-55.
- Thompson, Jane, "The Ability of Children of Different Grade Levels to Generalize on Sorting," Journal of Psychology, XI, January, 1941, 119-126.

III. UNPUBLISHED MATERIALS

Barber, E. R., A Study of Scatter and the Relative Difficulty of Sub-tests in the Revised Stanford-Binet, Unpublished Master's Thesis, University of Illinois, Champaign, Illinois, 1938.

Rau, Frances Virginia, The Performance of Young Children in a Classification Task, Unpublished Master's Thesis, Loyola University, Chicago, Illinois, 1943.

APPROVAL SHEET

The thesis submitted by Sister M. Fidelissima Dzik, Fel., O.S.F. 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 5, 1952
Date

M. Lucella A. Twomey
Signature of Adviser