Loyola University Chicago
Loyola eCommons

Master's Theses | Theses and Dissertations

1966

The Otis as Intelligence Test: A Comparison of 2nd and 4th Grade Scores

Virginia Wenzel Hill
Loyola University Chicago

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

Part of the Psychology Commons

Recommended Citation

Hill, Virginia Wenzel, "The Otis as Intelligence Test: A Comparison of 2nd and 4th Grade Scores" (1966). *Master's Theses*. 2141.
https://ecommons.luc.edu/luc_theses/2141

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 © 1966 Virginia Wenzel Hill

THE OTIS As INTELLIGENCE TEST:

A COMPARISON OF 2nd AND 4th GRADE SCORES

BY

Virginia Wenzel Hill

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

January

1966

ACKNOWLEDGMENTS

The author wishes to express her indebtedness to Charles I. Doyle, S. J., and Thomas M. Kennedy, Ph. D., for their support and assistance in directing this thesis.

Appreciation is likewise due to the author's husband, James, whose encouragement and aid were vital to the completion of this manuscript.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	Introduction	1
II.	Review of the Literature	4
III.	Procedure and Presentation of the Data	14
IV.	Discussion	21
V.	Summary and Conclusions	26
BIBLIOGRAPHY		28

LIST OF TABLES

TABLE PAGE

1. Scores on Otis As Group Intelligence Test 15

LIST OF FIGURES

FIGURE PAGE

1. Comparison of Second and Fourth Grade Scores 16

2. Comparison of Second and Fourth Grade Scores: Low Ability Group 18

3. Comparison of Second and Fourth Grade Scores: High Ability Group 19

4. Comparison of Second and Fourth Grade Scores: Middle Ability Group 20

CHAPTER I

INTRODUCTION

The assessment of intellectual ability has been of concern to man from the earliest times. With the inception of psychological testing educators and clinicians formulated the problem in more definite terms. The French physician Esquirol, as early as the middle of the nineteenth century, tried various procedures to classify the degrees of feeblemindedness. The term "mental test" was first used by James McKeen Cattell in an article in 1890 in which he described his results in testing college students. From these early beginnings a variety of intelligence tests have been developed, both for group and individual administration.

Today's extensive use of psychological tests has come under criticism for alleged inaccuracies and for the stereotypic effect it may have on children. Much misunderstanding of psychological tests, especially intelligence tests, is generated by misuse and misinterpretation of these tests. It is generally agreed by test advocates that the most accurate administration and interpretation of intelligence tests is made when tests are given on an individual basis by trained personnel.

In our school setting today, however, individual tests are not feasible because of the limitations of time and personnel. By necessity then, group intelligence tests are heavily relied upon to indicate the ability of the

school child. Certain limitations are inherent in group tests, including factors within the test, the child, or the environment. Johnson (1954) sums these up as: factors of the test including reliability and validity of the test, appropriateness and accuracy of the norms, type of test, administration, scoring, standard deviation and standard error of the test; factors within the individual such as emotional or organic defects of the child; and finally these factors combined where the score is in the environment but affects the child, these include cultural differences, language, experience, age and sex. Being aware of the varied factors that may influence a score on an intelligence test, other than the actual ability of the person, we become more skeptical of considering test scores apart from further study of the child and the test. "Standardized tests must be discussed in terms of the whole child with full realization that the test mark may very well be incorrect. Our conclusions, then, should be tentative and not a permanent brand to be worn by the child as he moves through the school system." (Ricken, 1958)

Test scores, particularly IQ scores are most helpful in understanding a child if they are properly used. However, if they are used in an uninformed or too narrow manner, they can hinder understanding by stopping further investigation or help. They can be used ascclues to understanding, but not as a key to understanding a child.

With these factors in mind, the investigator proposes to study the accuracy and applicability of one of the many group intelligence tests. In this study the Otis Quick-Scoring Mental Ability test, Alpha short form (Otis As) will be investigated. This particular aspect of the wide field of

intelligence testing was chosen for two reasons: 1) the parish schools in the Chicago Archdiocese use this test extensively; 2) in a recent thesis Grant (1961) compared two group intelligence tests with the Stanford-Binet individual intelligence test and found that while both tests differed significantly from the Stanford-Binet, the Otis did so slightly more. Because of these findings it was felt that more detailed analysis of the Otis test would be useful.

In this study the Otis As intelligence scores for a group of fourth graders will be compared with their second grade scores to determine the constancy of this measure. The group will be broken down into low ranking, middle ranking and high ranking youngsters to study if the ability level of the child affects the variability of the score. This work studies the test as used in the classroom with average children. We will attempt to ascertain if the factor of the ability level of the child affects the stability of his score on this test. Marked changes in the fourth grade score from the second grade score would indicate that one or both are in error.

CHAPTER II

REVIEW OF THE LITERATURE

It is hoped that this study will lead to a deeper understanding of the strengths and weaknesses of the Otis Quick-Scoring Mental Ability Tests short form. The Alpha Test is for grades 1-4, the Beta Test for grades 4-9 and the Gamma Test for high schools and colleges. The Alpha Test: Short Form is an abbreviated version of the original Alpha test. While the original test yields separate verbal and nonverbal IQ scores, the As measures both, but combines them into a single score. The content of this test is completely pictorial or geometric. The test purports to "measure mental ability -- thinking power or the degree of maturity of the mind." (Otis, 1954) The nonverbal part of the test requires that the student select the different picture from each series, the basis of differentiation varying. The same series of pictures is used for the verbal part but one picture is to be marked in each row in compliance with specific directions, hence measuring the youngsters ability to follow directions and make a variety of distinctions.

The author states that "the purpose of the new series is to provide 'self-administering' tests which can be scored more quickly than any of the preceding tests." (Otis, 1939) This test aptly meets its purpose in that it can be administered to a group in half an hour and is quickly scored with a key which enables the right answers to be counted without marking the items

right or wrong. Another feature of this test for lower grade use is that it does not require reading skills.

The nonverbal test of the original Otis was developed in a pragmatic manner being revised through three experimental editions using over 600 students. The verbal test was similarly tested to insure its diagnostic value. (Otis, 1939) This normative sample was taken from 30 communities of 16 states throughout the country. (Barnes, Personal communication, 1962) The Otis short Form was developed from the original Otis using a representative sample of 1600 third grade students who had taken this test. From the original 190 items 45 were chosen which were the most differential between the good and poor groups. The correlation between Otis Alpha Original and Short form was .95. (Otis, 1954)

The validity of the original Otis Quick-Scoring test was ascertained by computing the coefficients of correlation between the Alpha test and the Primary Examination, and between Alpha and Grade Placement. The r for the former was .65 and for the latter .86. The probable error of a score for the Total Score was 4.6 points. In this validity study for the original test second and third grade students were used. (Otis, 1939) The Alpha, Short Form, was tested for validity by measuring the correlation between the Alpha Short Form As and the average of two reading subtests of the Metropolitan Achievement Test. After being corrected for attenuation the coefficients for two independent samples were .69 and .68. The Alpha original scores (converted into As scores) were also correlated with various subtests of the Stanford Achievements testswith correlations ranging from .31 (Parragraph

meaning in first grade) to .63 (Arithmetic Reasoning in fourth grade). (Barnes, 1955) This correlation with achievement type of validity was chosen "because one of the major purposes for which a school uses an intelligence test such as Alpha is to provide a basis for estimating ability to handle school work successfully, it is proper to think of 'validity' as the extent to which the Alpha scores are related to and can be used to predict, school achievement." (Otis, 1954)

Two samples of third grade pupils (N = 370 in each sample) were used to compute Split-Half reliability coefficients for form As. Using the Spearman-Brown formula the coefficients were .87 and .88. The standard error of mreasurement is given in the manual as 4 score points. Since the relationship between score and IQ is not linear Table 2 in the manual which gives the IQs corresponding to deviations of score must be consulted. (Otis, 1954) "A 4 point score difference near the IQ = 145 level corresponds to 8 points of IQ; at the IQ = 125 level, to 7 IQ points; at IQ = 100 to 6 IQ points; at IQ = 85, to 5 IQ points; at IQ = 70 to 3 IQ points. (Summary)

Published studies on the Otis As are sparse, considering the wide use of this test. Before reviewing these studies a survey of the general literature on intellectual functioning and intelligence testing would be useful. For Wechsler (1958) "Intelligence, operationally defined, is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment." Goddard (1947) also defines intelligence in this adaptive, problem-solving frame of reference, but places more emphasis on its training aspects. To him it is "the degree

of availability of one's experiences for the solution of his present problems and the anticipation of future ones." A more abstract, theoretical concept of intelligence is given in B. L. Welman's statement that "I personally like to think of intelligence as quality of thinking, of making abstractions, generalizing, manipulating symbols, applying and adapting information and knowledge to new and different situations." (Welman, 1944)

From these quotations we can see that intelligence is an ability, but like most abilities it can only be measured in its application to concrete situations. Good (1954) points out that measuring intelligence is a form of inference: the performance is measured from which we first infer an ability and from this infer a capacity. From such inferences there are, of course, possibilities of errors at many points. If the observed performance is not the person's optimum, or if the demonstrated ability has not been properly developed, then any inference to capacity is incorrect. Good states these possibilities in a positive way when he comments that the quality of the performance of a child on an intelligence test is based on: 1) what he is capable of learning, 2) the opportunities he has had to learn, and 3) his willingness to give evidence of having learned. (Good, 1954) It is important to keep in mind that intelligence is an ability, but that the ability can be present but not demonstrated for one reason or another. Intelligence tests, then, measure the present functioning of the ability but cannot be interpreted to measure the limits of the ability.

As our school system becomes larger and more complex, psychological tests are increasingly relied upon to aid in making decisions. Great care

is needed to insure that these intended aids do not contribute to injustices and poor education. Too often the early labeling of children with IQ scores can become a self fulfilling prophesy in that the school system inadvertently structures the education so that the youngster lives up to his "measured" IQ. Benjamin Bloom, former chief of the examiner's office at the University of Chicago, points out that "we are in real danger of creating an IQ elite by singling as future leaders the small proportion of our youth who have special abilities needed to do well on standardized tests. We are screening out and throwing away other qualities we may need even more to guide us safely through the space age — not only character traits like enterprise and daring but basic mental abilities not even recognized by the tests." (Lagemann, 1962)

Since errors can easily be made with any IQ test, provision needs to be made for further study of individual children when possible problems are indicated. This intensive follow-up study, along with the proper understanding of the limitations of IQ scores will help to correct many of the misuses of these tests. When a low score is obtained on a test various aspects of the situation need to be discussed. Is the child physically fit? Is the social-economic and educational background similar to that of the normative sample? Is the youngster free from emotional disturbances? Has the test been properly administered and scored? What are the weaknesses of this test? Would a test score from a carefully administered individual test reveal a truer picture of the child's ability? "It should be stated categorically, perhaps, that no group test of any kind should be used unless

there be provision for intensive, individual study of those persons making low scores on it." (Worcester, 1947)

In the published literature there were no reports of studies similar to ours. In correspondence with the Test Department, Harcourt, Brace & World, Inc., Tarrytown, New York, results of studies they had made or which had been sent to them were obtained. The Otis As IQ scores were compared with scores from other tests:

from: 1957 Technical Report on the California Test of Mental Maturity
CTMM Language Factors r = .57
CTMM Language (Short Form) r = .52
CTMM Non-Language Factors r = .39
CTMM Non-Language (Short Form) r = .33 (California, 1957)

Westbury, L. I., New York School Year 1956-57
Second Grade 140 cases
Mean Otis Alpha IQ 116.54
Mean CTMM IQ 115.46
r = .673
Fourth Grade 54 cases
Mean Otis Alpha IQ 104.04
Mean CTMM IQ 114.31
r = .762

Lucas County (Ohio) Child Welfare Board 20 cases
Mean Otis Alpha IQ 67.1
Mean Stanford-Binet IQ 68.0
r = .44

Public School System of a large Southern City 128 cases
Mean Otis Alpha IQ 82.9
Mean Stanford-Binet IQ 85.4
r = .87

A Maine School System
First Grade 230 cases
Mean Otis Alpha IQ 98.1
Mean Stanford-Binet IQ 100.0
r = .53
Second Grade 220 cases
Mean Otis Alpha IQ 102.2
Mean Stanford-Binet IQ 100.6
r = .54

Third Grade 197 cases
Mean Otis Alpha IQ 98.2
Mean Stanford-Binet IQ 99.6
r = .64
Fourth Grade 150 cases
Mean Otis Alpha IQ 94.8
Mean Stanford-Binet IQ 98.0
r = .53

(Summary)

In an unpublished thesis (Grant, 1963) the validity of the Kuhlmann-Anderson group intelligence test and the Otis Quick-Scoring Mental Ability tests were studied using the Revised Stanford-Binet, Form L as the validating criteria. The data were taken from tests of children seen at Loyola Center for Guidance. To check the accuracy of Loyola's test findings he compared them to individual IQ test scores administered to the same children elsewhere (r = .91). The group test scores were then correlated with the individual IQ test scores. The correlation between the Kuhlmann-Anderson and the Stanford-Binet (r = .78) was slightly higher than that between the Otis Alpha and the Stanford-Binet (r = .74). Both group tests tended to rate the IQs lower than the individual tests did. This group was not broken into ability levels, hence these findings are for the group as a whole.

Mildred M. Allen (1945) conducted a study similar to the present study in the New Rochell, New York Public Schools. A comparison was made of the Kuhlmann-Anderson Intelligence Test scores administered midway through first grade and at the beginning of fourth grade. The relationship was given between the Mental Age, IQ and percent of average development (Kuhlmann prefers this measure over IQ, determined by dividing an individual's mental unit points by the average mental unit points for his age group). The MA in

grade IV is as equally related to either MA, IQ or PcAv obtained in grade I with correlation coefficients of .51, .51 and .52. Only a moderate degree of relationship is found as correlations of this size have corresponding coefficients of alienation (k) of .8602, .8602, and .8542, and indices of forecasting efficiency of E = 13.92, 13.92 and 14.58, respectively. Thus the error in predicting the Kuhlmann-Anderson MA in Grade IV from any of these measures in Grade I is about 14% less than it would be if no correlation existed between the original and subsequent measures. The prediction of the fourth grade IQ and PcAv from first grade MA is no better. The correlation between grade I MA and grade IV IQ is .43 with an E value of 9.72 and k=.9028. The fourth grade PcAv correlated .39 with the first grade MA. The E for an r of .39 is 7.92 (k = .9208) indicating a reduction of error of prediction of almost 8%. The author suggests that this low relationship may be due to the difference in content material of the two batteries (ie. inclusion of verbal tests in fourth grade) and/or inconstancy in mental growth. Allen concludes that the predictive value of intelligence tests given at an early age is highly questionable because of the non-verbal nature of these tests. With respect to the prediction of the IQ and PcAv in fourth grade from the same measures obtained in first grade, the predictive efficiency is about three times as high as when made from Grade I MA, but these still do not predict fourth grade scores with much accuracy. For the correlation of .69 a k of .7238 is yielded and E = 27.62.

The New York City Board of Education conducted a study to determine the efficiency of the Otis Quick-Scoring As as a devise for possible screening

of candidates for classes for Children with Retarded Mental Development (IQ less than 75). The Otis As was administered to all of the third grade students in May, 1950. Of those tested 868 received an IQ score of 76 or below. The Stanford-Binet, L was then administered to this group. Otis Mean IQ equaled 68.90, SD = 5.72; Stanford-Binet IQ Mean equaled 77.37, SD = 8.18. The correlation between the two was .42. While this correlation was too low to provide accurate prediction it was found that 62% of the pupils with an Otis IQ of 64 or below were eligible for CRMD classes. Hence the Otis As with a cut off at 64 IQ can be a time saving devise for screening CRMD candidates. (Slutsky, Justman and Wrightstone, 1952)

The original Otis Quick Scoring test was praised by Cattell for its ease and speed of administration and scoring. He pointed out, however, that the coefficients for reliability were low and that he considered the data for test norms inadequate. (Cattell, 1938)

Kuhlmann feels that while the test items are ingenious and exceptionally well chosen they lack variety and thus reduce in number the abilities measured. The test purports to measure general mental ability but there are too few items to adequately test the range. The "nonverbal" test is limited to the recognition of similarities and the "verbal" test is equally limited. Interest would not be sustained because the same type of questions are repeated. (Kuhlmann, 1949)

In commenting upon the shortened As version Alfred Yates, Research Officer, London, England, states that the test is useful to ascertain coarse scholastic classification when relevant information about the child's

previous educational progress is not obtainable. He feels that measures of attainment and the teacher's judgement are equally serviceable for the normal purposes of classification and guidance within a school. (Yates, 1959)

D. A. Lefever, Professor of Education at the University of Southern California, comments that the statements in the manual are vague and that the definition of the normative sample is not clear and that few facts support the author's statement that the norms should be thought of as representative of the country as a whole. He does agree, however, that "the Otis Quick-Scoring Mental Ability Tests, as the title implies, do furnish a short easily scored indicator of scholastic aptitude. Such a measure, if interpreted with care, can be useful to both teacher and counselor by revealing within fairly broad limits of accuracy the probable level of academic achievement for a majority of pupils." (Lefever, 1959)

CHAPTER III

PROCEDURE AND PRESENTATION OF THE DATA

In order to understand more fully the constancy of the IQ score when measured by the Otis As intelligence test the present study compares the scores of over 700 students for two administrations of this test. These students were drawn from ten parish schools of various socio-economic levels. The testing was done by the school personnel, generally the teacher administering the test in the classroom. (These are the usual circumstances under which this test is administered.) The scores from the various schools were combined, yielding an n of 713. A t test was used to determine if there is a significant difference between the second and fourth grade scores for this group.

The individual scores are presented and graphically compared in Table 1 and Figure 1, the second grade M = 107.600, SD = 14.351; the fourth grade M = 103.250, SD = 11.418. The Pearson Product Moment correlation for the two groups was r = .602. The Standard Error of Difference between the two means was .437. Applying a t test to this difference, the t of 9.93 is significant at the .001 level of confidence.

In order to see if the stability of this IQ measure was effected by the ability level of the student, the group was divided into three levels based on the second grade scores. The low ability group consisted of those scores

TABLE 1

SCORES ON OTIS As GROUP INTELLIGENCE TEST

Intelligence Quotients	Second Grade Frequency	Fourth Grade Frequency
148-152	5	1
143-147	5	0
138-142	8	0
133-137	26	1
128-132	28	5
123-127	30	15
118-122	53	57
113-117	85	66
108-112	106	124
103-107	123	137
98-102	78	86
93-97	81	99
88-92	42	72
83-87	26	31
78-82	18	20
73-77	7	7
68-72	2	2
63-67	0	0
Total	723	723
Mean IQ	107.600	103.250
SD	14.351	11.018

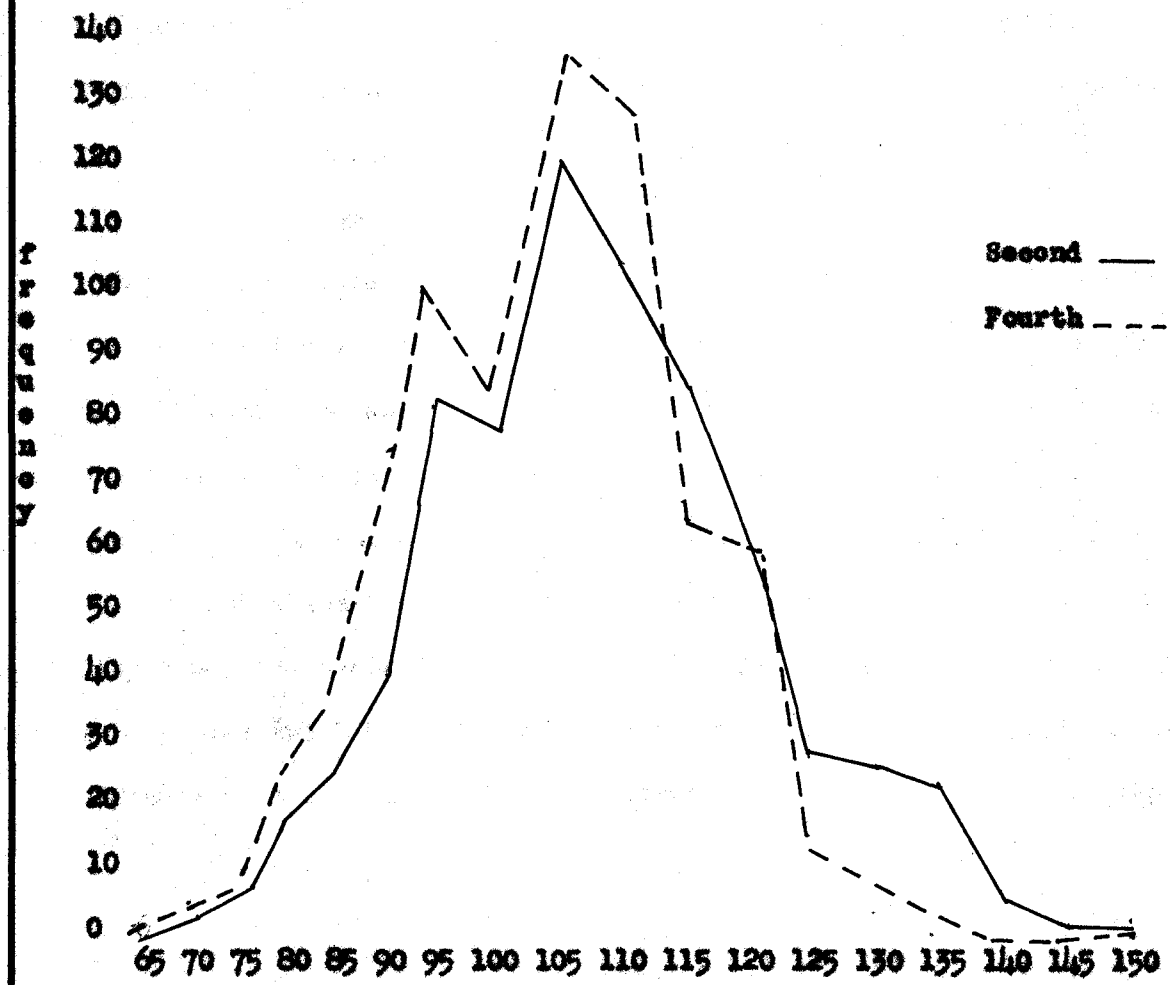

FIGURE 1

COMPARISON OF SECOND AND FOURTH GRADE SCORES

less than IQ 92. This is approximately 1 SD below the mean. The high ability group consisted of those scoring above IQ 123, which is approximately 1 SD above the mean. The middle group consisted of the scores from IQ 93 to IQ 122. A comparison was then made between the second and fourth grade scores of these groups. (Figs. 2, 3, 4)

For the low ability group the Mean IQ score in second grade was 85.37. SD = 5.193; in fourth grade the mean was 91.68, SD = 9.830. The difference of 6.31 IQ points between these two means is significant at the .001 level of confidence. For the high ability group the mean of 132.3039, SD = 6.846 dropped to 114.754, SD = 7.902 in fourth grade. This difference of 18 IQ points is also significant at the .001 level of confidence. For the middle ability group the second grade mean was 106.853, SD = 2.438, and the fourth grade mean was 103.127, SD = 9.954. Again the difference of 3.727 IQ points is significant at the .001 level of confidence.

FIGURE 2

COMPARISON OF SECOND AND FOURTH GRADE SCORES

Low Ability Group

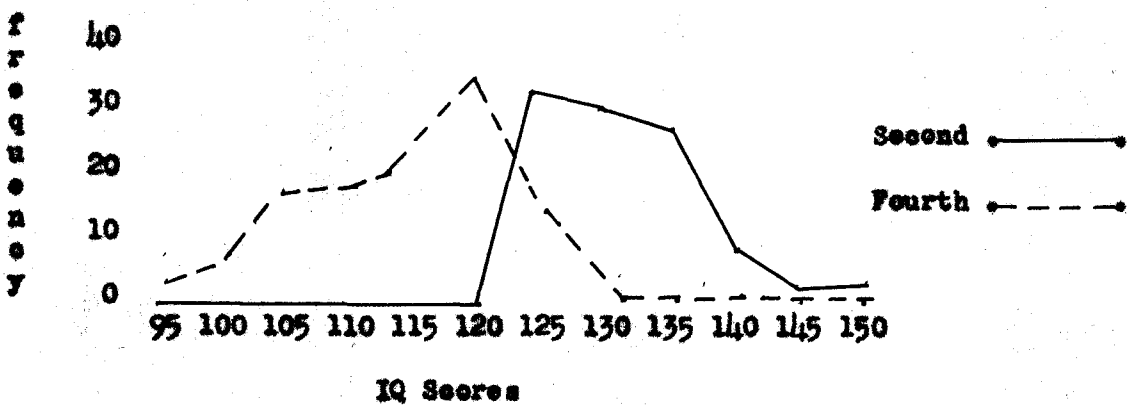

FIGURE 3

COMPARISON OF SECOND AND FOURTH GRADE SCORES

High Ability Group

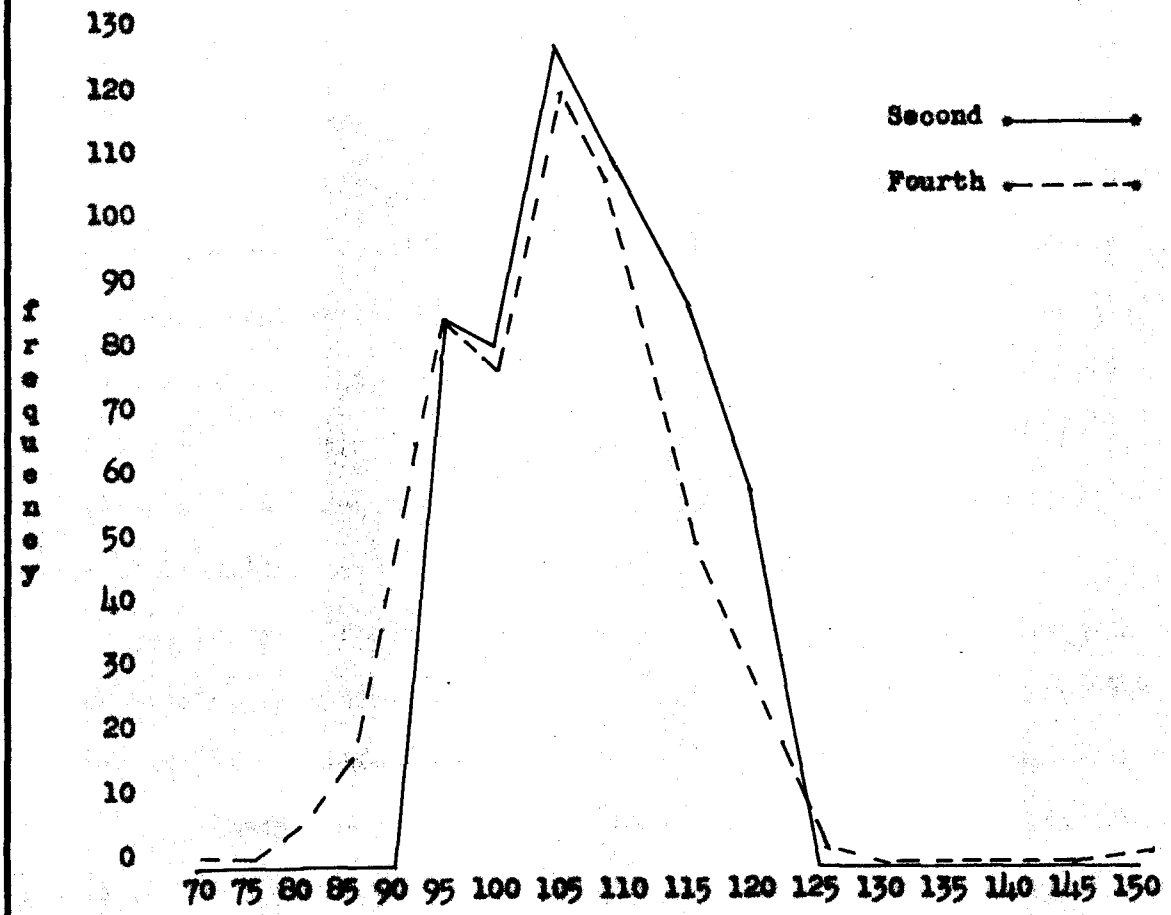

FIGURE 4

COMPARISON OF SECOND AND FOURTH GRADE SCORES

Middle Ability Group

CHAPTER IV

DISCUSSION

The present study measures the long range stability of the Otis As. "Such information belongs more properly under the heading of validity than under the heading of reliability, since it concerns broad, enduring behavioral changes rather than temporary fluctuations in specific test performance." (Anastasi, 1961) As evidenced by various studies, retest correlations are higher when the interval between tests is shorter and the children are older at the time of testing. Using the Stanford-Binet Intelligence test a correlation of r = .83 was obtained between tests given at three and four years of age. Correlations between this original three year test and tests at successive ages decreased steadily until at age 12 the correlation had dropped to r = .46. (Sontag, 1958) A similar study (Bradway, 1949) shows the test-retest correlation to be r = .59 when a group of preschool children were retested 25 years after the original administration of the test. A correlation of r = .65 existed after a 10 year interval. This same group tested at 14 years of age and retested 15 years later resulted in a correlation of r = .85. Thus indicating that at an older age test-retest correlations are higher.

In the light of these correlations and the high esteem in which the

Stanford-Binet is held in its use as a model in validating other tests, our test-retest correlation of .60 suggests the justifiability of using the Otis As when a short group test is needed for the general assessment of intelligence. The relative ease with which the Otis As can be administered and scored and the results which compare favorably (for a group) to the tool which is often considered the most accurate intelligence test for school age children, make it reasonable to use the Otis As as a tool for measuring intellectual development.

While our correlation suggests this conclusion the significant difference between the means of the two grade levels warrants further questioning before such a conclusion is accepted. Because a study of the scores indicated trends at the low and high ability levels and to analyze the strengths and weaknesses of this test for particular ability levels, it was decided to classify the subjects according to the criteria of ability level.

For the low ability group--those whose IQ measured less than 92 in second grade--there was a rise of a little over 6 IQ points in the means from second to fourth grade, which is significant at the .001 level of confidence. While several facts might explain this we feel that a consideration of the average mental age of this group in second grade offers some understanding both of the children and the test. Assuming, for the purpose of discussion that the fourth grade mean IQ is more accurate than the second grade mean IQ, we can determine a hypothetical mental age for the group in second grade from this higher IQ score. Such a mental age would be 6-10. While this might be the ability of these children they were not able to

demonstrate such ability at this time because a mental age of 6-10 is too young to perform adequately on a group, pencil-paper IQ test. (The actual measured mental age for the low group in second grade was 6-4) Hence, these youngsters of somewhat limited ability appear to be further below average when tested at an early age. By fourth grade they have developed sufficient mental maturity to perform on the group test.

For the middle ability group the difference between the mean IQ score is 3.73 points. While this is a statistically significant drop there does not seem to be a trend to explain the drop. Since this drop is less than the standard error of the score it is not large enough to be concerned about in the practical order.

The high ability group--those whose IQ measured greater than 122 in second grade--showed a mean IQ drop of 18 points from second to fourth grade. Not only is this drop significant at the .001 level of confidence but it is also the largest mean change of the three ability levels. This fall in mean IQ is not what is expected for this age group. Such a fall suggests that either the first administration over-rated the group or the second administration underrated them. In the light of parallel studies it seems reasonable to conclude that this form of the Otis Intelligence test is inadequate to measure intelligence at this ability level, and age.

Two similar studies lend weight to this conclusion. The Berkeley Growth Study of children from one month through eighteen years included tests of the youngsters at the same ages as our testing was done. The IQs for the children in this study tend to be well above the average. The Stanford-Binet 1916 revision was administered to 46 subjects at 7 years of age. The

mean IQ was 123.0 with a SD = 15.1. Two years later 45 of this sample were tested using the Stanford-Binet form L, with a resulting mean IQ of 129.0 and SD = 22.2. (Bayley, 1949) This group, both because of its superior intelligence and its age at testing is comparable to our high ability group. The 1937 revision of the Stanford-Binet yielded a higher score for the Berkeley group than did the 1916 revision. The 1916 and 1937 Stanford-Binet scores were also compared for a superior group by Ebert (1941) and the 1937 scores were found to be higher. However, he did find a consistent tendency for the means of the 1937 revision to increase with age from 6 to 10 years as did the Berkeley study. The author of that study concludes that the rise in mean IQ is a function of the age at which the tests were administered as well as the change in testing tools. It is of importance to note that in these two studies using the Stanford-Binet Intelligence Test the mean IQs rise at these ages while in our study there is a significant drop in mean IQ. This reinforces our original conclusion of an inadequacy in the As for this ability level and age.

Factors that may account for this suggested inadequacy in the Otis As may possibly be found in the test itself. For this form of the Otis series the highest Mental Age obtainable is 12-11 for a performance in which all of the items are completed correctly. Such a "perfect" performance by the average age child in fourth grade would yield an "Alpha" IQ of 139. In theory then, the ceiling on this form is adequate for the brighter fourth grade students. However, in practice it may be questionable because even the brightest subject will often make errors in a test situation, not due

to lack of ability but rather due to extraneous factors such as inattention, misunderstanding of specific directions, nervousness, etc. This then would possibly bring the ceiling of the test to a level that would not adequately tap the limits of some children. Another factor in the test performance is the subject's attitude. The form given in fourth grade is the same as the one given in second grade. By necessity, then, the format is at a young interest level. The older, brighter students may possibly react in a somewhat negative manner to this test, both because of its format and because they had taken the test previously. This negative attitude could result in their not putting forth their whole effort to perform on the test.

These areas: 1) the immaturity of lower ability children to participate in a group, pencil-paper IQ test in second grade; 2) too low a ceiling on the Otis As to adequately test the brighter fourth grade students and 3) the brighter students' reaction to the format of this test, offer material for further study.

CHAPTER V

SUMMARY AND CONCLUSIONS

The present study is concerned with the long range stability of a widely used intelligence test for elementary schools. The Otis Alpha is the first in the Otis Quick Scoring series of intelligence tests and is often administered in second and fourth grades. Our study compares the scores obtained on two such administrations for 723 subjects from ten parish schools.

The mean IQ in second grade was 107.6000, SD = 14.351; and in fourth grade the mean was 103.25, SD = 11.41. The difference between these means is significant at the .001 level of confidence. The correlation of .602 is comparable to other test-retest correlations.

To further determine the strengths and weaknesses of the test for particular ability levels the group was divided into three ability levels based on second grade scores. There was a significant difference between the means for all three levels. For the low ability group there was a 6 IQ point rise from second to fourth grades. For the high ability group there was an 18 point drop. It was hypothesized from these findings that the lower ability youngsters may not be sufficiently mature to perform adequately on a group, pencil-paper IQ test in second grade. Moreover, for the brighter fourth grade students it seemed that the ceiling was too low to adequately test their

ability and that a test they had taken in second grade might fail to hold their interest when taken in fourth grade.

From our study the following conclusions seem justifiable:

1) Further study is needed to clarify the question raised by our findings regarding the ability of some second grade children to perform adequately on the Otis As.

2) For the brighter students this form of the Otis Quick scoring test is a weak tool in fourth grade. The test publishers recommend that in this grade either the Alpha or the Betta be given. The former is preferred at times because no reading is required. However, our results suggest that the Betta is the preferred form if the youngsters reading level is adequate.

BIBLIOGRAPHY

Allen, Mildred M. Relationship between the indices of intelligence derived from the Kuhlmann-Anderson intelligence tests for grade I and the same tests for grade IV. J. educ. Psychol, 1945, 36, 252-256.

Anastasi, Anne. The concept of validity in the interpretation of test scores. Educ. psychol. Measmt, 1950, 10, 67-78

_____ Psychological tests: uses & abuses. Teachers College Record, 1961, 62, 389-393.

Barnes, P. J. Prediction of achievement in grades 1 through 4 from Otis-Quick Scoring Mental Ability Tests: Alpha Short Form. Educ. Psychol. Measmt, 1955, 15, 493-494.

_____ Personal communication, September 27, 1962.

Bayley, N. Consistency and variability in the growth of intelligence from birth to eighteen years. J. genetic Psychol, 1949, 75, 165-196.

Bonney, M. E. Relative stability of social, intellectual and academic status in grades II to IV and the inter-relationships between these various forms of growth. J. educ. Psychol, 1943, 34, 88-102.

Bradway, K. P., Thompson, C. W. & Crawens, R. B. Preschool IQs after 25 years, J. educ. Psychol., 1958, 49, 278-281.

Cattell, P. B. in O. K. Buros (Ed.), The nineteen thirty-eight mental measurements year book, New Brunswick: Rutgers U. Press, 1938.

Charles, C. M. A comparison of the intelligence quotients of three different mental tests applied to a group of incarcerated delinquent boys, J. appl. Psychol, 1933, 17, 581-584.

Cronbach, L. J. Test reliability: its meaning and determination, Psychometrika, 1947, 12, 1-16.

DeLong, A. R. Emotional effects of elementary school testing. Und. Child, 1955, 24, 103-107.

DeLong, A. R. Meaning of individual scores on group tests. National council on measurements used in education, 14th year book, 1957, 43-49.

Dumas, S. S. Critical survey of methods of computing and using IQs. British J. ed. Psychol, 1956, 26, 137-140.

Ebert, E. H. A comparison of the original and revised Stanford-Binet Scales. J. Psychol, 1941, 11, 47-61.

Findley, W. G. Group testing program for the modern school. Ed. Psychol. Measmt, 1945, 2, 173-179.

Goddard, H. H. Suggested definition of intelligence. Train sch. Bull. 1947, 43, 185-193.

Good, W. R. Misconceptions about intelligence testing. Ed. Digest, 1954, 20, 14-16.

Grant, J. G. A comparison of the revised Stanford-Binet, form L with the Kuhlmann-Anderson test and the Otis Quick-Scoring Mental Ability Tests, (unpublished thesis) Chicago: Loyola University, 1961.

Guilford, J. P. New standards for test evaluation. Ed. psychol. Measmt, 1946, 6, 427-438.

_____ The structure of intellect. Psychol. Bull, 1956, 53, 267-293.

Johnson, G. B., Jr. Factors to be considered in the interpretation of intelligence-test scores. Elemt. school J, 1954, 145-150.

Kuder, F. in O. K. Buros (Ed.), The third mental measurements year book, New Brunswick: Rutgers U. Press, 1949.

Kuhlmann, F. in O. K. Buros (Ed.). The nineteen-forty mental measurements year book, Highland Park, New Jersey, 1940.

Lagemann, J. K. Let's abolish IQ tests. Ed. Digest. 1962, 27, 1-4.

Lefever, D. A. in O. K. Buros (Ed.), The fifth mental measurements year book, Highland Park, New Jersey, 1959.

Lighthall, F., Ruebush, B., Sarason, S. & Zweibelson, I. Change in mental ability as a function of test anxiety and type of mental test. J. consltng. Psychol, 1959, 23, 34-38.

Lincoln, E. A. & Wadleigh, V. L. Change and constancy in group IQs. School & Society. 1932, 35, 606-608.

McRae, H. Inconstancy of group test IQs. British J. ed Psychol, 1942, 12, 59-70.

Manwiller, C. E. Constancy of the IQ based on group measurement. Pittsburgh School, 1937, 11, 125-143.

Phillips, A. J. Comparison of methods of calculating mental age equivalents. J. educ. Psychol, 1943, 34, 152-165.

Pidgeon, D. A. Design, construction and use of standardized tests and the interpretation of test scores. Educ. Research. 1961, 4, 33-43.

Ricken, R. Misuse of standardized tests of intelligence and scholastic aptitude. High Points, 1958, 40, 61-62.

Skaggs, E. B. Some critical comments on certain prevailing concepts used in mental testing. J. appl. Psychol, 1927, 11, 503-508.

Slutsky, J. E., Justman, J. & Wrightstone, J. W. The use of a group intelligence test as a screening device for the selection of mentally retarded children for placement in special classes. Amer. J. mental Deficiency, 1952, 57, 106-108.

Sontag, L. W., Baker, C. T. & Nelson, Virginia L. Mental growth and personality development: a longitudinal study. Monogr. Soc. Res. Child Developmt, 1958, 23, #2.

Stroud, J. B. Intelligence tests in school use. J. educ. Psychol, 1957, 48, 77-86.

Summary of the reliability data for certain group intelligence tests with special reference to the Otis Quick-Scoring Mental Ability tests. Harcourt, Brace and World, Test department, Tarrytown, New York.

Technical report on the California Test of Mental Maturity, California, 1957.

Terman, L. M. The intelligence of school children, New York: Houghton Mifflin

Wechsler, D. Measurement and appraisal of adult intelligence, Baltimore: The Williams & Wilkins Co., 1958.

Welman, Beth L. Some misconceptions about intelligence. Childhood Educ, 1944, 21, 108-112.

Worcester, D. A. Misuse of group tests in the schools. Educ. psychol. Measmt, M7, No. 4, 779-781.

Yates, A. in O. K. Buros (Ed.), The fifth mental measurements year book. Highland Park, New Jersey, 1959.

Zweibelson, I. Test anxiety and intelligence test performance. J. consult. Psychol, 1956, 20, 479-481.

APPROVAL SHEET

The thesis submitted by Virginia Wenzel Hill 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.

January 18, 1966
Date

Thomas M. Fennell
Signature of Adviser