1963

An Investigation of the Effects of Item Placement on Test Performance

Thomas John Ginley
Loyola University Chicago

AN INVESTIGATION OF THE EFFECTS OF ITEM PLACEMENT ON TEST PERFORMANCE

by

Thomas John Ginley

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

1963

LIFE

Thomas John Ginley was born in Chicago, Illinois, on January 25, 1938.

He was graduated from Loyola Academy in June, 1956 and Loyola University, Chicago, Illinois, June, 1960 with the degree of Bachelor of Science. He began his graduate studies in the Department of Psychology at Loyola University in June, 1960.

The author was employed as a member of the test construction division of the Chicago Civil Service Commission, December, 1960 to February, 1963. He is presently employed as Director, Division of Educational Measurements, American Dental Association.

ACKNOWLEDGEMENTS

The writer wishes to express his gratitude to Dr. Edmund P. Marx, Professor in the Department of Psychology, Loyola University for his encouragement and advice in the preparation of this study.

To Dr. Robert Nicolay and Dr. Ronald Walker are due most sincere thanks for patience, understanding and assistance in the securing of subjects for this project.

The author also wishes to thank the Psychological Corporation for the permission to reprint the Concept Mastery Test of Dr. Lewis Terman.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

"No other contribution of psychology has had the social impact equal to that created by the psychological test" (Guilford, 1954). Since the advent of the original age scales developed by Binet for the study of intelligence, psychological testing has transgressed into a highly diversified and virtually overaccepted position.

The basic function of psychological testing is to measure differences between individuals or between performances of the same individual under different circumstances. The initial application of psychological testing emphasized the identification of intellectual deficiency and currently remains a prominent use of specific types of psychological tests. Educational problems provided a basis for additional areas of test development. The desire to classify students with reference to their ability to profit from different types of school instruction, the diagnosis of academic failures, counseling of high school and college students, the selection of aspirants for professional schools are a few of the educational uses of psychological testing (Anastasi, 1961).

The selection and placement of business and industrial personnel together with the certification of employees under civil service represent a recent and rapidly expanding use of psychological testing.

An examination of the latest edition of the Mental Measurements Yearbook (Buros, 1959) will attest to the rapid growth within the test development field.

On the basis of this trend, there is evidence supporting an additionally significant increase during the next decade.

The major exploitations of psychological testing have emphasized prediction and assessment. For example, it is highly desirable to discuss selection devices, personality inventories, intelligence testing, aptitude test batteries and an array of similar topics within the context of ordinary conversation. The possible explanation for this pronounced interest in testing arises chiefly from the competitive nature of our present social climate. Parents currently "need" to know the I.Q of their child, his educational outlook as compared with others, and whatever additional assessments are available.

The college graduate of today will have encountered entrance examinations, aptitude test batteries, numerous achievement tests and various other mental assessments during his educational process and possibly even a "Wonderlic or Otis" while applying for a job.

This growth of psychological testing therefore necessitates a keener awareness of the associated problems in addition to full acceptance of the responsibilities concurrently attached. The mere fact that testing has attained its current position of acceptance does not of itself indicate that such status is objectively warranted. Rather, it typifies the need for sound stratified research to empirically substantiate the basis of current test theory.

Therefore, the purpose of this study is to investigate one element of basic test construction theory - the placement of items.

It has been the consensus of test construction researchers since the advent of testing and the Binet scale that the optimum arrangement of test items should proceed in the order of increasing difficulty. Although this practice

has gained almost universal acceptance, the rationale for it seems to be based on "accepted logic" rather than experimental evidence. A careful search of the literature has failed to disclose a comprehensive study adequately investigating this problem.

The subsequent chapters of this research project will deal specifically with an investigation of item placement from the standpoint of significant differences in performance.

CHAPTER II

REVIEW OF LITERATURE

The early investigators in the field of mental measurement - Galton (1883), Cattell (1888), Kraepelin (1895), Ebbinghaus (1897) had focused their efforts on the study of individual differences from the standpoint of sensory and perceptual processes.

Binet and Henri (1895) criticised most of the available test series as being overly sensory in nature and therefore concentrating unduly on simple, specialised abilities. They argued that, in the measurement of the more complex functions, precision is not necessary, since individual differences are larger in these functions. They proposed a varied list of tests covering such functions as memory, imagination, attention, comprehension, suggestibility, aesthetic appreciation, and many others (Anastasi, 1961).

Binet published, "L' Etude Experimentale de L' Intelligence" (1903) in which he subjectively investigated his two daughters, Armande and Marguerite, on their ability to perform 20 given tests. From investigations such as this into the analysis of various aspects of intelligence, the Binet - Simon Scale emerged.

The Binet - Simon Scale (1905) contained 30 tests which were arranged in ascending order of difficulty. The difficulty level of the tasks was imperially determined by administering the tests to 50 normal children aged 3 to 11 years, and to some intellectually deficient children. The tests were designed to

cover a wide variety of intellectual functions. They were tests of intelligence but in 1905, Binet had only a vague idea of what he meant by intelligence (Varon, 1935). Binet scored the tests by adding age increments for each successfully accomplished task. In the Terman revisions, the number of tasks increased but the process remained essentially the same. The items indicated that the child who passed a test successfully possessed an ability that ". . . corresponds to the average ability of children of such and such an age" (Terman, 1937).

It was through this method of scoring, namely, that of adding increments from unrelated scores and thereby obtaining a single total based on many diverse tasks, that Binet and subsequently Terman avoided the problem of absolute scaling. Currently, this problem is frequently overlooked since it is customary to use Binet I.Q. results as total scores when assessing, for example, the placement of children within school programs. Terman was aware of this problem and recommended the use of standard scores rather than I.Q.'s to indicate performance. Terman's faith in the adherence of his suggestions is apparent in his statement (Terman, 1937), "Whatever index of brightness is used, some will claim too much from it and others too little. The uninformed will read meaning into it which it does not connote and the over enthusiastic will, in too exclusive dependence upon it, ignore their lines of information which should be taken into account." As a result, Terman suggests that simple indexes be used and that training into their significance and limitations be adequately given.

From this brief overview of individual intelligence testing, and its costly procedures, it becomes somewhat apparent why the shift toward group

testing emerged. It also is logical that any transition from individual to group testing would include as many of the acceptable procedural conditions of individual testing as possible. However, is this method of item arrangement, namely, the ascending order of difficulty, a necessary procedural condition to all good test construction? Subsequent chapters will hopefully present some information on this subject.

Proponents, such as Ruch (1929) state, "The easiest item should come first and the most difficult last with all intervening items arranged in order of increasing difficulty." This statement is presented with several other "rules" of test construction, however, no rationale is mentioned by Ruch as the basis for this specific request. Ross (1947) also suggests, "The items in the test should be arranged in order of difficulty. It is especially important to have the easiest items at the beginning and the hardest ones at the end of the test. The exact order of the intervening items is less important." Patterson (1925) recommends that several initial items be made so easy that all subjects will pass them. This recommendation arises from the "shock absorber" theory of testing, which holds that the initial test in a series be sufficiently easy that the person is able to perform on it without any effort, thereby starting the actual test series without any initial shock. It is Patterson's feeling that this type of arrangement will diminish the candidate's apprehension and therefore permit a more stable transition into the task to be performed. Terman (1916), Cronbach (1946) and Adkins (1947) and others make similar recommendations without any supporting evidence.

Several authors indicate acceptance of this method from the standpoint of "minimal disturbance." Ross (1947) indicates, "One of the problems of

measurements is to arrange conditions so that the thing being measured is disturbed as little as possible in the act of measuring." Traxler (1951) suggests the complexity of the problem by noting his observation that intellectual behavior is distorted by any disruption of the mental functioning of the subject. It is his feeling that items should carefully be arranged according to difficulty in order to preserve the most desirable motivating circumstances, which he defines as, ". . . those which enable the largest number of individuals to turn in the best performance without undue emotional stress." Traxler further suggests that tension may increase performance in some subjects, "Others become nervous and over-excited in a test situation and thus may find their performance blocked by their emotional state."

Capron (1933) investigated the effects of presenting items by varying the order of difficulty. Three types of achievement tests were used in this research project, arithmetic problems, spelling, and fundamental processes in arithmetic. The item presentation of easy-to-hard, hard-to-easy, and random order was given to a small group of subjects. The results demonstrated no significant difference in performance on any of the presentations. Capron concluded there was no empirical basis for the claim that items must proceed from easy-to-hard as has been accepted.

A note of caution must be inserted while discussing Capron's study. The number of subjects was small by current standards and the controls within the experimental design left much to be desired. It is therefore suggested that the results be analyzed within a somewhat hypercritical frame of reference.

Lund (1953) conducted a similar investigation and attempted to relate test performance to order of item difficulty, anxiety and intelligence. The major

emphasis of this study concerned itself with anxiety and test performance as interpreted through the use of the Taylor Anxiety Scale and the Henmon - Nelson Test of Mental Ability.

The Henmon - Nelson test was varied with relation to item placement. Form A remained the conventional arrangement of items while Form B utilized the last 25% of the items in Form A as its first 18 items with the remainder unchanged. "Schematically represented, the two forms consist of 72 identical items with the order varied as indicated:

		Hard Items
Form A	1. 54	55. 72
	Hard Items	
Form B	55 72	1. 54

The items in Form B are renumbered to avoid revealing the item rearrangement to the subjects."

Two matched groups of 90 subjects each were administered the two forms of the test. The same subjects were divided into three groups with respect to their performance on the Taylor Anxiety Scale. The conclusions drawn by Lund indicated that a change of item order had a significant effect on test performance and that this effect was negative at all levels of intelligence. There was, however, no significant difference reported between anxiety level and test performance. Lund's overall inference suggests that, ". . . the standard practice of arranging test items in the conventional progressive order of difficulty be continued in order to provide a consistent optimum level of motivation among the subjects."

CHAPTER III

PROCEDURE

The present study was undertaken to investigate the effect that order of item presentation has on test performance. It was intended to discover whether or not there was a significant difference between performance on tests in which items were arranged in the conventional easy-to-hard order as contrasted with performance on tests in which a reversal of this order was presented.

In order to fulfill the requirements of an adequate experimental design, the author found it necessary to use a standardized, single factor, power test on which item analysis data were available - as an integral part of this study, Part I of the Concept Mastery Test - Form T by Dr. Lewis M. Terman was selected.

The Concept Mastery Test - is a measure of ability to deal with abstract ideas at a high level. The test consists of two parts: I, the identification of synonyms and antonyms, and II, the completion of analogies. The items have been so selected as to draw on concepts from a wide variety of subject matter fields, such as physical and biological sciences, mathematics, history, geography literature, music and so forth.

The Concept Mastery Test was devised at Stanford University in 1939 for use in the follow-up program of the Stanford research with gifted subjects. After a survey of the results yielded by the leading types of intelligence tests - their reliabilities, their validities as measures of "intellect" and their

relative efficiency per unit of time - two types of tests were chosen: the synonyms - antonyms and the analogies tests. The test was given its present name because it deals chiefly with abstract ideas.

The Concept Mastery Test is designed for group administration and is easily scored on a IBM 805 Test Scoring machine. There is no time limit since the Concept Mastery Test is a test of power rather than speed.

Part I (synonyms - antonyms) of the Concept Mastery Test consisted of 115 items arranged in the order of increasing difficulty:

ITEM

1. rigid . . . . . flexible
2. competent . . . . . qualified

.

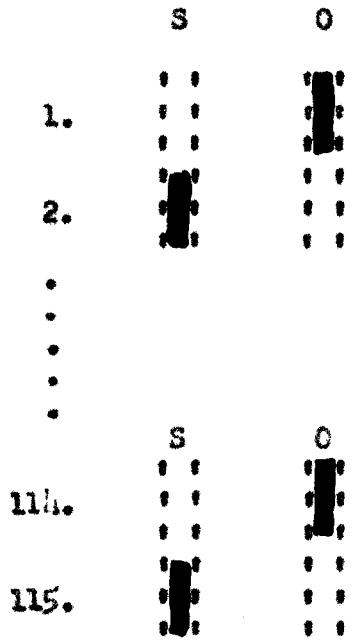
.

.

.

114. disingenuous. . . . . artless
115. transilient . . . . . saltatory

The subjects are asked that if the two words have the same or nearly the same meaning to blanken the space under S on the answer sheet or if they mean the opposite or nearly the opposite to blanken the space under O.



On the basis of the intended experimental design, the following procedural methods were undertaken:

1) Part I was divided in half in order to establish two parallel forms of the test.

Form A included items

1-3-5-.....-109-111-113

of the original test in ascending order of difficulty

Form B included items

2-4-6.....-110-112-114

of the original test in ascending order of difficulty.

Both Forms consisted of 57 items. Item number 115 of the original scale was omitted.

The items were renumbered consecutively to avoid any confusion in administration and to conceal the item rearrangement.

2) Form A was then reconstructed in reverse order thereby establishing Form C.

3) Form B was also reconstructed in reverse order thereby establishing Form D.

Forms C and D are therefore presented in the order of decreasing difficulty.

4) The items on Form A were randomly rearranged by means of table of random numbers. The result was the establishment of Form E.

5) The items on Form B were also randomly rearranged by means of a table of random numbers. The result was the establishment of Form F.

Six experimental groups were established to investigate the specific effects of the item rearrangement on these six sub-forms of the Concept Mastery Test. Each of the groups consisted of 42 subjects with an overall total N of 252. The 252 subjects were male and female undergraduate students that had not completed their sophomore year of college at the time of the test administration.

Table I shows the experimental procedure used in the administration of the tests to the six groups of subjects. Each subject was tested twice and therefore became his own experimental and control index.

TABLE I

ORDER OF TEST ADMINISTRATION

	1st Administration	2nd Administration
Group I N = 42	Form A	Form D
Group II N = 42	Form B	Form C
Group III N = 42	Form C	Form B
Group IV N = 42	Form D	Form A
Group V N = 42	Form A	Form F
Group VI N = 42	Form B	Form E

Prior to the administration of the six forms under the imposed experimental conditions, a pilot study was conducted to determine if Forms A and B were indeed, parallel tests. A statistical criterion for parallel tests was observed (Gulliksen, 1950) and the means, variances, and intercorrelation were investigated.

CHAPTER IV

RESULTS

Since the assumption of parallel tests is crucial to this experiment, a preliminary study was conducted to investigate whether or not Form A and Form B met the statistical criterion for parallel tests (Gulliksen, 1950).

A group of 42 subjects, male and female freshmen and sophomore undergraduate college students were administered test Forms A and B. The order of test presentation was counterbalanced:

TABLE II

PILOT STUDY - ORDER OF TEST PRESENTATION

N = 42	N	1st Administration	2nd Administration
Pilot	21	A	B
Group	21	B	A

The results of the preliminary investigation to ascertain the parallel nature of tests Forms A and B are incorporated in Table III.

TABLE III

COMPARISON OF GROUP TEST PERFORMANCE ON FORM A AND FORM B

N= 42	FORM A	FORM B
Mean	33.10	34.93
σ	8.90	10.30
σ^2	79.21	106.09
σ_M	1.39	1.60

σ_{M_D} - 1.735

r - .621

C.R. - 1.05 *

F - 1.33 *

*NOT SIGNIFICANT AT .05 LEVEL OF CONFIDENCE.

The results in Table III show that neither the critical ratio of 1.05 nor the F ratio of 1.33 is significant at the 5% level of confidence, therefore, the two tests can be considered sufficiently parallel for this investigation.

The 252 subjects contained in the six groups of this experimental study were given a set of instructions concerning the use of the Concept Mastery Test. Among additional information, the subjects were told not to guess and to work only in the order in which the items appear. Since this test has only two possible answers for each question /same (S) or opposite (O)/, the scoring formula suggested by Terman is "Rights minus Wrongs." This author initially scored the answer sheets by means of the uncorrected raw score (Total rights).

However, it became increasingly apparent that guessing was a predominant factor in the administration of the test so it was therefore necessary to employ Terman's correction scoring formula. On the basis of the corrected scoring method, it became necessary to add a constant of 20 points to all 252 papers in order to bring the negative scores to zero or above. A Pearson product moment correlation was computed comparing the raw score method (Rights) and the adjusted scoring method (Rights minus wrong plus 20) and yielded a coefficient of .754 with an N = 504.

The specific results comparing test performance of each group appears within Tables IV, V, VI, VII, VIII, and IX.

TABLE IV

GROUP I

COMPARISON OF GROUP TEST PERFORMANCE ON FORM A AND FORM D

N = 42	FORM A 1st Administration	FORM D 2nd Administration
Mean	32.20	33.00
σ	6.80	7.20
σ^2	46.24	51.84
σ_M	1.03	1.12

σ_{M_D} - 1.079

r - .498

C.R. - .74

TABLE V

GROUP II

COMPARISON OF GROUP TEST PERFORMANCE ON FORM B AND FORM C

N = 42	FORM B 1st Administration	FORM C 2nd Administration
Mean	35.55	32.71
σ	11.60	9.00
σ^2	134.56	81.00
σ_M	1.81	1.40

σ_{M_D} - 1.37

r .662

C.R. 2.07 *

*SIGNIFICANT AT THE .05 LEVEL OF CONFIDENCE.

TABLE VI

COMPARISON OF GROUP TEST PERFORMANCE ON FORM C AND FORM B

N = 42	FORM C 1st Administration	FORM B 2nd Administration
Mean	34.60	34.25
σ	11.60	10.60
σ^2	134.56	112.36
σ_M	1.81	1.65

σ_{M_D} - 1.38

r - .694

C.R. - .257

TABLE VII

COMPARISON OF GROUP TEST PERFORMANCE ON FORM D AND FORM A

N = 42	FORM D 1st Administration	FORM A 2nd Administration
Mean	35.80	34.80
σ	8.60	10.25
σ^2	73.96	105.06
σ_M	1.34	1.60

σ_{M_D} - 1.42

r - .687

C.R. - .845

TABLE VIII

GROUP V

COMPARISON OF TEST PERFORMANCE ON FORM A AND FORM F

N = 42	FORM A 1st Administration	FORM F 2nd Administration
Mean	32.70	36.29
σ	9.79	10.44
σ^2	94.83	108.99
σ_M	1.53	1.63

σ_{M_D} - 1.42

r - .59

C.R. - 2.45 *

*SIGNIFICANT BEYOND THE .05 LEVEL OF CONFIDENCE.

TABLE IX

GROUP VI

COMPARISON OF GROUP TEST PERFORMANCE ON FORM B AND FORM E

N = 42	FORM B 1st Administration	FORM E 2nd Administration
Mean	36.00	36.60
σ	9.70	10.60
σ^2	94.09	112.36
σ_M	1.51	1.65

σ_{M_D} - 1.28

r - .673

C.R. - .469

The results indicated in Table V show a Critical Ratio of 2.07 which is significant at .05 level of confidence. The mean performance of Form B was 35.55 as compared with 32.71 on Form C. On the basis of this information alone, a statistically significant difference in performance seems to exist.

The results of Table VIII show the only other Critical Ratio which typifies a significant difference in performance, 2.45 which is significant beyond the .05 level of confidence. Form F has a mean of 36.29 as compared with 32.70 on Form A. Again, a real difference in performance seems to exist. If all of the experimental conditions were adequately met, this difference should be attributable to the order of item presentation. However, a closer look at the results tends to negate a real significant difference in performance but rather indicates a flaw inherent in test Form B. A detailed analysis of the means will tend to support this conclusion.

TABLE X

MEAN PERFORMANCES OF THE SIX EXPERIMENTAL GROUPS

	FORM	FORM
Group I	A	D
Mean	32.20	33.00*
Group II	B	C
Mean	35.55*	32.71
Group III	C	B
Mean	34.60	34.25*
Group IV	D	A
Mean	35.80*	34.60
Group V	A	F
Mean	32.70	36.29*
Group VI	B	E
Mean	36.00*	36.60

_____ The mean performances on Form A and its derivatives.

* The mean performances on Form B and its derivatives.

Upon inspection of Table X, Form B and its derivates seem to produce a mean slightly higher than that of Form A. This could suggest that Form B was slightly easier for this group of subjects. If this is the case, and the evidence tends to point in that direction, the significant differences noted in Tables V and VIII are at least partly attributable to the ease of Form B and not solely to the order of item presentation.

On the basis of the entire set of results, evidence tends to indicate that there was no real effect on test performance by varying the order of item presentation other than the extraneous results caused by inadequate controls.

There is in addition, a hint that the random order of item presentation, on a vocabulary test such as the Concept Mastery Test will produce a slightly improved performance over the other methods of item arrangement. Group V and Group VI tend to substantiate this inference. In Group V, the subjects significantly improved on their performance on the random test form. (Note: Form F is a derivative of Form B and therefore easier). Group VI also gave some indication of this improvement by reporting a mean on Form E, slightly higher than that of Form B (the easier test). If this is a real difference, a possible explanation for the improvement on the random test forms could be attributed to additional reading time incurred on the part of the subjects.

In conclusion, the results of this experiment generally indicate that item rearrangement on a vocabulary test such as the Concept Mastery Test, has no consistent significant effect on individual test performance.

CHAPTER V

SUMMARY AND CONCLUSIONS

The standard practice in test construction is to arrange the items of the test so that the easier items are placed first and that the remaining items gradually increase in difficulty with the more difficult items placed last. Although there has been general agreement that this is a "rule" of good test construction, a careful search of the literature has not revealed an adequate experimental study which directly investigates this problem.

This present study was undertaken to investigate the effect that order of item presentation has on test performance. It was intended to discover whether or not there would be a significant difference between performance on tests in which items were arranged in the conventional easy-to-hard order as contrasted with performance on test in which a reversal of this order or no specific order at all was presented.

Six experimental forms were established from Part I of the Concept Mastery Test by Lewis M. Terman. Two parallel forms of each experimental condition were reprinted (2 forms easy-to-hard, 2 forms hard-to-easy, 2 random forms).

Six experimental groups were established to investigate the specific effects of the item rearrangement on these sub-forms of the Concept Mastery Test. Each group consisted of 42 college subjects. Each subject was tested twice, therefore, a measure of his performance would be secured under the

normal administration in addition to the specific experimental treatment.

The results were analyzed in terms of group significant differences. It was discovered that Form B was slightly easier than Form A and therefore partly responsible for the significant differences among two groups.

The overall conclusion that should be drawn on the basis of the existing results is that item placement, in terms of difficulty level, on a verbal test such as the Concept Mastery Test has no consistant significant effects on individual test performance.

BIBLIOGRAPHY

Adkins, D. C. Construction and analysis of written tests for predicting job performance. Educ. Psychol. Measmt., 1946, 6, 195-211.

Adkins, D. C. Construction and Analysis of Achievement Tests. Washington, D. C.: United States Civil Service Commission, 1947.

Anastasi, Anne. Psychological Testing. New York: The Macmillan Co., 1961.

Anderson, J. E. The effect of item analysis upon the discrimination power of an examination. J. Appl. Psychol., 1935, 19, 237-244.

Binet, A. L' Etude Experimental de L' Intelligence. Paris: Costes, 1903.

Binet, A., and Simon, Th. Methodes nouvelles pour le diagnostic du niveau intellectual dis anorwaux. Annee Psychol., 1905, 11, 191-244.

Boring, Edwin G. A History of Experimental Psychology. New York: Appleton-Century-Crofts, 1957.

Buros, Oscar. The Fifth Mental Measurements Yearbook. New Jersey: The Gryphon Press, 1959.

Capron, Virginia. The relative effect of three orders of arrangements of items upon pupil scores in certain arithmetic and spelling tests. Journal of Educational Psychology, 1933, 24, 687-695.

Carroll, J. B. The effect of difficulty and chance success on correlations between items or between tests. Psychometrika, 1945, 10, 1-19.

Cronbach, Lee. Essentials of Psychological Testing. New York: Harper and Bros., 1960.

Cronbach, L. J. Response sets and test validity. Educ. Psychol. Measmt., 1946, 6, 475-494.

Galton, F. Inquiries into Human Faculty and its Development. London: Macmillan, 1883.

Goodenough, Florence L. Mental Testing: Its History, Principles, and Applications. New York: Holt, Rinehart and Winston, 1949.

Guilford, J. P. Psychometric Methods. New York: McGraw-Hill, 1954.

Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw Hill, 1950.

Gulliksen, H. Theory of Mental Tests. New York: Wiley, 1950.

Lund, K. E. Test Performance as Related to Order of Item Difficulty, Anxiety, and Intelligence. Unpublished Ph.D. Dissertation, Northwestern, 1953.

Mollenkopf, W. G. An experimental study of the effects on item analysis data of changing item placement and test time limit. Psychometrika, 1950, 15, 291-315.

Nunnally, J. C. Tests and Measurements. New York: McGraw-Hill, 1959.

Paterson, Donald C. Preparation and Use of New-Type Examinations. New York: World Book, 1925.

Remmers, H. H. Educational Measurement of Evaluation. New York: Harper and Bros., 1943.

Richards, T. N. Modern Clinical Psychology. New York: McGraw-Hill, 1946.

Ross, C. D. Measurement in Today's Schools. New York.

Ruch, S. M. The objective or new-type examination. An Introduction to Educational Measurement. Chicago: Scott, Foresman and Company, 1929.

Terman, L. W. The Measurement of Intelligence. Boston: Houghton Mifflin Co., 1916.

Terman, L. W. and Merrill, M. Measuring Intelligence. Boston: Houghton Mifflin Co., 1937.

Traxler, A. E. Administering and scoring the objective test. In Lindquist, E. F., Educational Measurement. Washington: American Council on Education, 1951.

APPENDIX I

CONCEPT MASTERY TEST

Lewis M. Terman

Your answers are to be recorded on a special answer sheet. Please do not make any marks on this test blank. Fill out the heading on the answer sheet before starting the test.

SYNONYMS AND ANTONYMS

Directions:
This test is made up of pairs of words which have either the same or opposite meaning. If two words mean the SAME or nearly the same, mark the space under "A" on the answer sheet; if two words mean the OPPOSITE or nearly the opposite, mark the space under "B" on the answer sheet.

For example:
In item 101 below, "hot" and "cold" mean the opposite, so that the space under "B" should be blackened.

In item 102 below, "big" and "large" mean the same, so that the space under "A" should be blackened.

101.	hot.....cold	101.	A B C D " ▌ " " " ▌ " "
102.	big.....large	102.	A B C D ▌ " " " ▌ " " "

Be sure that the space you mark on the answer sheet is numbered the same as the question you are answering. DO NOT MARK THE TEST BLANK. Omit those items that you could answer only by pure guess, but answer all you THINK you know, even if you are not quite certain. Do not study long over any pair.

WORK ONLY IN THE ORDER IN WHICH THE QUESTIONS APPEAR

FOR RESEARCH ONLY

1. rigid.....flexible
2. haven.....refuge
3.
4.
5.

6.
7.
8.
9.
10.

11.
12.
13.
14.
15.

16.
17.
18.
19.
20.

21.
22.
23.
24.
25.

26.
27.
28.
29.
30.

31.
32.
33.
34.
35.

36.
37.
38.
39.
40.

41.
42.
43.
44.
45.

46.
47.
48.
49.
50.

51.
52.
53.
54.
55.

56. abstruse.....recondite
57. imprecate.....execrate

Form B

1. competent.....qualified
2. cheap.....priceless
3.
4.
5.

6.
7.
8.
9.
10.

11.
12.
13.
14.
15.

16.
17.
18.
19.
20.

21.
22.
23.
24.
25.

26.
27.
28.
29.
30.

31.
32.
33.
34.
35.

36.
37.
38.
39.
40.

41.
42.
43.
44.
45.

46.
47.
48.
49.
50.

51.
52.
53.
54.
55.

56. termagant.....virago
57. disingemuous.....artless

1. imprecate.....execrate
2. abstruse.....recondite
3.
4.
5.

6.
7.
8.
9.
10.

11.
12.
13.
14.
15.

16.
17.
18.
19.
20.

21.
22.
23.
24.
25.

26.
27.
28.
29.
30.

31.
32.
33.
34.
35.

36.
37.
38.
39.
40.

41.
42.
43.
44.
45.

46.
47.
48.
49.
50.

51.
52.
53.
54.
55.

56. haven.....refuge
57. rigid.....flexible

1. disingenuous.....artless
2. termagant.....virago
3.
4.
5.

6.
7.
8.
9.
10.

11.
12.
13.
14.
15.

16.
17.
18.
19.
20.

21.
22.
23.
24.
25.

26.
27.
28.
29.
30.

31.
32.
33.
34.
35.

36.
37.
38.
39.
40.

41.
42.
43.
44.
45.

46.
47.
48.
49.
50.

51.
52.
53.
54.
55.

56. cheap.....priceless
57. competent.....qualified

Form E

1. argot.....jargon
2. cursory.....desultory
3.
4.
5.

6.
7.
8.
9.
10.

11.
12.
13.
14.
15.

16.
17.
18.
19.
20.

21.
22.
23.
24.
25.

26.
27.
28.
29.
30.

31.
32.
33.
34.
35.

36.
37.
38.
39.
40.

41.
42.
43.
44.
45.

46.
47.
48.
49.
50.

51.
52.
53.
54.
55.

56. confute.....refute
57. expert.....tyro

Form F

1. bucolic.....rustic
2. spoliation.....despoliation
3.
4.
5.

6.
7.
8.
9.
10.

11.
12.
13.
14.
15.

16.
17.
18.
19.
20.

21.
22.
23.
24.
25.

26.
27.
28.
29.
30.

31.
32.
33.
34.
35.

36.
37.
38.
39.
40.

41.
42.
43.
44.
45.

46.
47.
48.
49.
50.

51.
52.
53.
54.
55.

56. dynamic.....static
57. descant.....expatiate

APPENDIX II

APPENDIX II

ITEM ANALYSIS DATA

Group I

Item Number	FORM A Number Correct	FORM A Proportion Correct	FORM D Number Correct	FORM D Proportion Correct
1	41	976	6	143
2	39	929	4	095
3	38	905	6	143
4	34	809	13	309
5	39	929	7	167
6	40	952	15	357
7	24	571	10	238
8	23	548	14	333
9	39	929	16	381
10	16	381	7	167
11	27	643	13	309
12	34	809	20	476
13	26	619	15	357
14	32	762	15	357
15	33	786	20	476
16	24	571	22	524
17	11	262	10	238
18	24	571	11	262
19	40	952	11	262
20	17	401	8	190
21	20	476	17	405
22	15	357	15	357
23	31	738	22	524
24	25	595	17	405
25	16	381	10	238
26	13	310	16	381
27	42	1.000	22	524
28	24	571	10	238
29	10	238	21	500
30	26	619	14	333
31	21	500	23	548
32	10	238	14	333
33	16	381	16	381
34	15	357	13	309
35	21	500	9	214

Continued...

ITEM ANALYSIS DATA

Group I (Continued)

Item Number	FORM A Number Correct	FORM A Proportion Correct	FORM D Number Correct	FORM D Proportion Correct
36	15	357	18	429
37	21	500	27	643
38	10	238	18	429
39	17	404	32	762
40	18	429	28	667
41	9	214	24	571
42	13	309	19	453
43	11	262	33	785
44	12	286	22	524
45	8	190	23	548
46	29	690	24	571
47	13	309	22	524
48	10	238	32	762
49	12	286	16	381
50	20	476	21	500
51	10	238	37	881
52	6	143	42	100
53	9	214	40	952
54	8	190	38	905
55	12	286	41	976
56	8	190	39	929
57	3	071	42	100

ITEM ANALYSIS DATA

Group II

Item Number	FORM B Number Correct	FORM B Proportion Correct	FORM C Number Correct	FORM C Proportion Correct
1	42	100	8	190
2	39	929	16	381
3	41	976	10	238
4	39	929	22	529
5	39	929	20	476
6	39	929	15	357
7	36	857	11	262
8	27	643	20	476
9	28	666	17	405
10	37	881	10	238
11	32	762	15	357
12	26	619	28	666
13	22	524	17	405
14	24	571	14	333
15	35	833	14	333
16	23	548	9	214
17	25	595	12	286
18	26	619	17	405
19	39	929	8	190
20	17	405	13	310
21	26	619	26	619
22	22	529	18	429
23	18	429	26	619
24	15	357	21	500
25	16	381	14	333
26	17	405	20	476
27	20	476	42	100
28	17	405	27	643
29	28	666	19	452
30	16	381	28	666
31	27	643	16	381
32	22	529	17	405
33	16	381	17	405
34	20	476	30	714
35	21	500	24	571
36	15	357	23	548
37	13	310	16	381
38	16	381	23	548

Continued...

ITEM ANALYSIS DATA

Group II (Continued)

Item Number	FORM B Number Correct	FORM B Proportion Correct	FORM C Number Correct	FORM C Proportion Correct
39	15	357	36	857
40	15	357	33	786
41	15	357	23	548
42	21	500	28	666
43	20	476	35	833
44	15	357	30	714
45	14	333	33	786
46	17	405	35	833
47	9	214	32	762
48	13	310	22	524
49	13	310	36	856
50	12	286	24	571
51	15	357	25	595
52	10	238	38	905
53	8	190	37	881
54	12	286	34	809
55	11	262	39	929
56	12	285	38	905
57	5	119	40	952

ITEM ANALYSIS DATA

Group III

Item Number	FORM B Number Correct	FORM B Proportion Correct	FORM C Number Correct	FORM C Proportion Correct
1	42	100	10	238
2	39	929	13	310
3	41	976	19	452
4	39	929	20	476
5	39	929	20	476
6	39	929	15	351
7	36	857	18	429
8	27	643	21	500
9	28	666	13	310
10	37	881	18	429
11	32	762	19	452
12	26	619	33	786
13	22	524	23	548
14	24	571	20	476
15	35	833	14	333
16	23	548	17	405
17	25	595	17	405
18	26	619	19	452
19	39	929	24	571
20	17	405	17	405
21	26	619	21	500
22	22	524	15	351
23	18	429	24	571
24	15	357	23	548
25	16	381	16	381
26	17	405	24	571
27	20	476	42	100
28	17	405	23	548
29	28	666	23	548
30	16	381	18	429
31	27	643	24	571
32	22	524	23	548
33	16	381	25	595
34	20	476	35	833
35	21	500	28	666
36	15	357	25	595
37	13	310	22	524
38	16	381	25	595

Continued...

ITEM ANALYSIS DATA

Group III (Continued)

Item Number	FORM B Number Correct	FORM B Proportion Correct	FORM C Number Correct	FORM C Proportion Correct
39	15	357	38	905
40	15	357	32	762
41	15	351	29	690
42	21	500	33	786
43	20	476	33	786
44	15	351	30	714
45	14	333	30	714
46	17	405	36	857
47	9	214	32	762
48	13	310	32	762
49	13	310	38	905
50	12	286	28	666
51	15	357	31	738
52	10	238	36	857
53	8	190	38	905
54	12	286	33	786
55	11	262	37	881
56	12	286	41	976
57	5	119	41	976

ITEM ANALYSIS DATA

Group IV

	FORM D		FORM A	
Item Number	Number Correct	Proportion Correct	Number Correct	Proportion Correct
1	8	190	42	100
2	7	167	40	952
3	14	333	41	976
4	15	357	32	762
5	13	309	39	929
6	8	190	40	952
7	13	309	29	691
8	13	309	28	666
9	19	452	37	881
10	13	309	25	591
11	19	452	31	738
12	20	476	34	809
13	14	333	31	738
14	23	548	29	690
15	19	452	32	762
16	27	643	29	691
17	19	452	20	476
18	12	286	38	905
19	24	571	39	929
20	18	429	28	666
21	13	309	9	214
22	19	452	24	571
23	24	571	34	810
24	25	595	24	571
25	19	452	14	333
26	23	548	9	214
27	32	762	42	100
28	11	262	20	476
29	20	476	14	333
30	21	500	19	452
31	19	452	19	452
32	14	333	15	357
33	19	452	19	452
34	15	357	14	333
35	22	524	21	500
36	24	571	12	286
37	29	691	17	405
38	23	548	11	262

Continued...

ITEM ANALYSIS DATA

Group IV (Continued)

	FORM D		FORM A	
Item Number	Number Correct	Proportion Correct	Number Correct	Proportion Correct
39	34	809	11	262
40	29	691	15	357
41	28	666	13	309
42	18	429	10	238
43	37	881	10	238
44	21	500	10	238
45	26	619	12	286
46	26	619	22	524
47	25	595	12	286
48	37	881	12	286
49	27	643	10	238
50	29	691	11	262
51	36	857	15	357
52	40	952	12	286
53	42	100	9	214
54	38	905	12	286
55	41	976	11	262
56	38	905	14	333
57	42	100	11	262

ITEM ANALYSIS DATA

Group V

	FORM A		FORM F	
Item Number	Number Correct	Proportion Correct	Number Correct	Proportion Correct
1	40	952	17	401
2	38	905	16	381
3	38	905	9	214
4	31	738	14	333
5	40	952	12	286
6	41	976	35	833
7	20	476	31	738
8	28	666	35	833
9	37	881	12	286
10	23	548	10	238
11	32	762	14	333
12	30	714	5	119
13	30	714	12	286
14	37	881	26	619
15	34	809	21	500
16	32	762	13	309
17	16	381	5	119
18	33	786	10	238
19	37	881	41	976
20	19	452	33	786
21	15	357	6	142
22	17	401	20	476
23	19	452	7	166
24	28	666	40	952
25	22	524	20	476
26	8	191	23	548
27	42	100	18	429
28	16	381	6	142
29	12	286	19	452
30	23	548	22	524
31	16	381	8	191
32	18	429	20	476
33	10	238	36	857
34	17	405	5	119
35	21	500	5	119
36	10	238	42	100
37	15	357	18	429
38	13	309	4	.095

Continued...

ITEM ANALYSIS DATA

Group V (Continued)

Item Number	FORM A Number Correct	FORM A Proportion Correct	FORM F Number Correct	FORM F Proportion Correct
39	10	238	41	976
40	14	333	28	666
41	13	309	19	452
42	10	238	15	357
43	10	238	15	357
44	10	238	20	476
45	9	214	18	429
46	24	571	13	309
47	9	214	11	262
48	8	190	11	262
49	11	262	11	262
50	17	405	18	429
51	8	190	5	119
52	8	190	12	286
53	13	309	16	381
54	10	238	10	238
55	13	309	18	429
56	11	262	40	952
57	5	119	12	286

ITEM ANALYSIS DATA

Group VI

Item Number	FORM B Number Correct	FORM B Proportion Correct	FORM E Number Correct	FORM E Proportion Correct
1	42	100	7	167
2	39	929	16	381
3	42	100	13	309
4	40	952	26	619
5	40	952	41	976
6	42	100	36	857
7	34	809	33	786
8	33	786	41	976
9	28	666	13	309
10	35	834	10	238
11	23	548	42	100
12	28	666	13	309
13	20	476	8	190
14	19	452	24	571
15	34	810	42	100
16	23	548	20	476
17	24	571	36	857
18	25	595	31	738
19	40	952	15	357
20	12	286	34	809
21	31	738	9	214
22	24	571	14	333
23	18	429	14	333
24	15	357	14	333
25	17	405	21	500
26	17	405	32	762
27	23	548	24	571
28	16	381	36	857
29	19	452	19	452
30	11	262	14	333
31	24	571	37	881
32	17	405	6	143
33	13	309	17	405
34	21	500	20	476
35	17	405	13	309
36	18	429	34	809
37	12	286	29	691
38	17	405	14	333
39	14	333	11	262

Continued...

ITEM ANALYSIS DATA

Group VI (Continued)

Item Number	FORM B Number Correct	FORM B Proportion Correct	FORM E Number Correct	FORM E Proportion Correct
40	14	333	8	191
41	8	190	19	452
42	28	666	16	381
43	16	381	21	500
44	12	286	18	429
45	18	429	8	191
46	15	357	12	286
47	10	238	28	666
48	7	167	12	286
49	14	333	29	691
50	11	262	9	214
51	10	238	13	310
52	9	214	31	738
53	9	214	33	786
54	11	262	20	476
55	9	214	39	929
56	5	119	11	262
57	8	191	10	238

APPROVAL SHEET

The thesis submitted by Thomas Joseph Ginley 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 10, 1963
Date

Edmund P. Marx
Signature of Adviser