

THE APPLICATION OF A SUBJECTIVE PROBABILITY TECHNIQUE
TO DIAGNOSTIC TESTS IN HIGH SCHOOL SUBJECT MATTER

A Thesis
Presented to
the Faculty of Graduate Studies
University of Manitoba

In Partial Fulfillment
of the Requirements for the Degree
Master of Education

by
Frederick Miller Olsen
1970



THESIS ABSTRACT

PROBLEM

The main purpose of this study was to examine the feasibility of using a subjective probability approach to testing in high school subjects. In the process, an attempt was made to estimate the relative merits of using the technique for (a) standardized tests, the results of which would be used as measures of aptitude (b) subject tests, the results of which could be used for predictive purposes and (c) diagnostic tests, the results of which would be used to expedite the review and reteaching of subject matter.

METHOD

The application of the subjective probability technique to scoring test items was assessed in four ways:

- (1) The results of standardized test items used in the subjective probability method were correlated with those of the same items treated in the conventional manner by the same students.
- (2) Specifically selected diagnostic test items were treated in both the conventional and the subjective probability manner by all students. These two sets of scores were also correlated.

- (3) The results of both conventional and subjective probability treatments of the diagnostic tests were correlated with mid-year examination results in the appropriate subjects.
- (4) Item analyses of each of the subject tests were performed to indicate the diagnostic information available from such a scoring technique.

It is generally the practice to postulate the "null hypothesis" and then, as a consequence of acquired information, to accept it, or to reject it, and accept an alternate hypothesis.

The following hypotheses were postulated:

- (1) The Spearman rank order correlation between the results of conventional and subjective probability treatments does not differ significantly from zero.
- (2) The Pearson product-moment correlation between conventionally treated subject tests and mid-term examination results does not differ significantly from the Pearson product-moment correlation between subjective probability treated subject tests and mid-term examination results.

The rejection of the null hypotheses would indicate that extra information might be made available from both ordinary standard test items and from items which had specifically been prepared. The subjective probability scoring technique could provide this information without destroying that available from the conventional results.

CONCLUSIONS

1. The subjective probability technique does not alter the rank of an individual from what it would be by using the conventional scoring method. This fact was true for both standardized tests and teacher-prepared diagnostic tests.

2. The subjective probability technique does not yield results which are of greater predictive value for essay-type, mid-term examinations than do the conventionally scored tests.

3. Partial knowledge can be readily assessed by means of analyses of frequency distributions of scores when the test items are scored by subjective probability techniques. This application could result in more efficient diagnostic and remedial work without altering any of the information concerning grading or prediction which is ordinarily available from multiple-choice tests.

ACKNOWLEDGEMENTS

This thesis was prepared under the direction of Dr. Peter A. Taylor, Associate Professor of Education at the University of Manitoba. He provided guidance and constructive criticism throughout the planning and writing of this study.

Thanks are also due Messrs. M. Cowie, G. Laidlaw and H. T. McCracken, who provided test items for the various diagnostic instruments and to Frank G. Garvin, Jr. who proof-read the text.

The author gratefully acknowledges these contributions.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	STATEMENT OF THE PROBLEM AND DEFINITION OF TERMS	1
	INTRODUCTION	1
	PURPOSE OF THE STUDY	4
	DEFINITION OF TERMS	6
	HYPOTHESES	7
	SUMMARY OF CHAPTER I	8
	OUTLINE OF THE PRESENTATION	9
II.	A REVIEW OF THE LITERATURE	10
	SUMMARY	18
III.	OUTLINE OF THE PROCEDURE	20
	INTRODUCTION	20
	PRELIMINARY ASSUMPTIONS	21
	TEST SUPPLIES	21
	SPECIFIC DIRECTIONS FOR ADMINISTRATION OF TESTS	22
	THE TESTING	24
	The Standardized Tests	24
	The Diagnostic Tests	25
	COLLECTION OF DATA	25
	TREATMENT OF DATA	28
	Scores	28
	Correlations	28

		vii
CHAPTER		PAGE
	Individual Item Tabulation	29
	SUMMARY	29
IV.	ANALYSIS OF DATA	31
	CORRELATION OF TEST DATA	31
	ANALYSIS OF CLASS KNOWLEDGE	43
	ANALYSIS OF INDIVIDUAL'S DEFICIENCIES	58
	STUDENTS' REACTIONS	66
	SUMMARY	66
V.	CONCLUSION AND EVALUATION	68
	BIBLIOGRAPHY	72
	APPENDIX	75

LIST OF TABLES

TABLE	PAGE
I. Raw Scores DAT-VR, Conventional then Subjective Probability, DAT-AR in Reverse Order	32
II. Raw Scores DAT-VR, Subjective Probability then Conventional, DAT-AR in Reverse Order	33
III. Spearman's Correlation to Identify Effects of Order of Scoring	34
IV. D.A.T. T-Scores	35
V. D.A.T. Percentile Ranks	36
VI. Grade XI I.Q. Scores	38
VII. Grade XII I.Q. Scores	39
VIII. Spearman's Correlation Relating I.Q.'s	40
IX. T-Scores of I.Q.'s	41
X. Spearman's Correlation Relating Results of Diagnostic Tests	42
XI. Comparison of Correlations Between Results of Diagnostic Tests and Related Mid-term Examinations	43
XII. Frequencies of Weightings of Items in Grade X Physics	45
XIII. Frequencies of Weighting of Items in Grade X Geography	47
XIV. Frequencies of Weighting of Items in Grade XI Geography	49
XV. Frequencies of Weighting of Items in Grade XI Chemistry	51
XVI. Frequencies of Weighting of Items in Grade XII Chemistry	53

TABLE

XVII.	Frequencies of Weighting of Items in Grade XII Physics	55
XVIII.	An Individual's Weighting of Items in Grade X Physics	60
XIX.	An Individual's Weighting of Items in Grade X Geography	61
XX.	An Individual's Weighting of Items in Grade XI Geography	62
XXI.	An Individual's Weighting of Items in Grade XI Chemistry	63
XXII.	An Individual's Weighting of Items in Grade XII Chemistry	64
XXIII.	An Individual's Weighting of Items in Grade XII Physics	65
XXIV.	Spearman's Rank Order Correlation Relating Scoring Techniques	69
XXV.	Fisher's Transformation Comparing Pearson Product Moments	69
XXVI.	Subjective Probability Scores for the Various Combinations of Weightings	110
XXVII.	Grade X Physics Results	112
XXVIII.	Grade X Geography Results	113
XXIX.	Grade XI Geography Results	114
XXX.	Grade XI Chemistry Results	114
XXXI.	Grade XII Chemistry Results	115
XXXII.	Grade XII Physics Results	116
XXXIII.	Grade X Physics	117
XXXIV.	Grade X Geography	119
XXXV.	Grade XI Geography	121
XXXVI.	Grade XI Chemistry	122

PAGE

TABLE

XXXVII.	Grade XII Chemistry	124
XXXVIII.	Grade XII Physics	126

CHAPTER I

STATEMENT OF THE PROBLEM AND DEFINITION OF TERMS

I. INTRODUCTION

For the past several decades there has been an increasing emphasis on the use of various types of discrete-answer objective tests. Some of the advantages of such tests when compared with essay-response tests are considered to be:

- (a) A wide sampling of skills and knowledge is possible;
- (b) Scoring is convenient and fast per unit;
- (c) Many adaptations of type of question are available;
- (d) High validity, reliability and usability are possible for most purposes.

The conventional arguments in favour of the essay-response have included:

- (a) The student learns to plan and express his argument coherently in essays;
- (b) It takes less time and effort to prepare an essay-response test;
- (c) For purposes of testing the specific techniques of writing, superior validity is achieved by the marking of the actual work.

The arguments in favour of the essay-response test

may be construed as criticisms of the objective testing technique. To the list of weaknesses in the objective-test program several others could be added. The conventional multiple-choice test tends to be designed, by the insertion of specifically planned distractors, to discover what the student does not know rather than to examine and diagnose what he has learned and understands. It is as though one could infer the store of information, ability and knowledge at hand by subtracting the unknown fraction from the sum total of knowledge available.

In many areas of study and discussion it is of importance that the students realize and accept that there are no "true" answers or "untrue" interpretations. When scored in the conventional way, the multiple-choice test item tends to eliminate this pedagogical or philosophical approach.

The chance of appreciably and consistently affecting a total score by random guessing has been conclusively refuted by R. F. Graesser (1966) as statistically insignificant. However, the fact remains that when one considers individual items of a test for analysis of responses, the probability of a lucky guess is possibly quite significant. One cannot be certain, when considering a "correct" response, whether it was:

- (a) a reflection of accurate knowledge of the information being examined;
- (b) the resultant of a combination of some knowledge and

a variety of subconscious clues (actually an infinite range of possible combinations); or

(c) merely a "lucky" guess reflecting no knowledge of the material in question.

Similarly, when an answer is one of those judged "incorrect", one cannot definitely infer that the response is due to a lack of knowledge or information. It may be that the "distractor" was selected because of a more complete knowledge of, or more perceptive thought about, the subject and a rejection of the trivial or superficial choice provided as the "true" answer.

To be sure, this latter type of item would be culled from a standardized battery because of the negative correlation which would be observed in detailed item analysis. With less formally constructed test items, incorrect conclusions may often result because of invalid assumptions. In order to discover information concerning the actual reason for a given response, personal interviews and discussion would have to be employed.

Referring specifically to those subject areas which lend themselves best to objective testing, it appears that it would be possible to extract additional information from the results of such tests without the associated disadvantages of the subjective items or the impractical expenditure of time involved in personal interviews. This procedure could be followed if details, concerning the form of and reason for the selection, were to be given by the

testee. An indication of confidence in, or agreement with, any or all of the possible selections would also have to be indicated by the subject as he answers the item. While the subjective probability technique basically produces a measure of the degree of confidence in a given topic, for practical purposes of testing, one could infer partial knowledge from the results. It is necessary that this process be both simple and objective for the student to make the indication and for the tester to assess it.

This greater measure of information could of course be put to use in more effective teaching or reteaching of subject areas. If it is to be a grading device, more accurate ranking or grouping could result because of the availability of a more complete measure of the knowledge of the student. If the test were to be designed for purposes of establishing entrance standards and hence ultimately used as a predictor, it should be possible to produce more reliable regression equations or prediction tables from the results because of the more accurate and complete picture afforded by the extra information. The degree of validity of such devices should correspondingly be increased.

II. PURPOSE OF THE STUDY

Because of the lack of diagnostic information available from the results of conventionally scored objective tests, it was decided to conduct an empirical investigation into the feasibility of the subjective pro-

bability approach to testing in high school. The technique was tested both with standardized test items and with items which had been prepared with a view to using the results as diagnostic tools for purposes of achieving identified aims in specific courses.

The application of the subjective probability method of marking test items was assessed in five ways:

1. The results of the items used from standardized tests which were selected for subjective probability treatment were correlated with those same items treated in the conventional method.
2. The results of the selected items of standardized tests treated in the subjective probability way were correlated with total scores in standardized tests.
3. The specifically prepared diagnostic test items were treated in both the conventional and the subjective probability manner and the results correlated.
4. The results of both conventional treatments and subjective probability treatments of the specially prepared subject tests were correlated with the mid-year examination results in the specified subjects.
5. Item analysis of each of the specially prepared test items was performed in order to indicate the extra diagnostic information available.

III. DEFINITION OF TERMS

Discrete Answer Test Item

This is an objective multiple-choice question. No provision for explanation or extenuation is provided.

Conventional Scoring Technique

This term refers specifically to discrete answer items in which only one response is judged of value for scoring purposes. The suggestion that partial knowledge may be present in the form of guessed answers has stimulated testers to correct for guessing in order to arrive at a "true" score. The formula used for the items in this study was:

$$\text{Score} = \text{number of correct responses} - \left(\frac{w}{n - 1} \right)$$

w = number of wrong responses

n = number of choices available per item

Students are informed that there will be a correction factor applied to nullify the effects of random guessing.

Subjective Probability Scoring Technique

This term also refers to multiple-choice questions but involves the selection of one or more of the provided alternatives as an answer. The subject is then expected to associate a numerical weight with each of the alternatives selected. The weights are to be in proportion to his confidence in the appropriateness of the choice and in this

study were limited to a sum total of five points per item. For practical purposes, this indication of degree of confidence has been equated with partial knowledge of the subject matter. Thus a weighting of five points on the one correct response receives maximum value. A distribution of the five points among several alternatives receives partial value for the item, while the indication of five points on an inappropriate choice receives zero credit. The formula used is the de Finetti equation which is outlined in the review of the literature.

IV. HYPOTHESES

The following null hypotheses are presented:

1. The Spearman rank order correlation between the results of conventional and subjective probability treatment does not differ significantly from zero.
2. The Pearson product-moment correlation between conventionally treated subject tests and mid-term examination results does not differ significantly from the Pearson product-moment correlation between subjective probability treated subject tests and mid-term examination results.

Because it has been the practice in classroom situations to consider T-scores, or rank in class, as a measure of a student's ability or knowledge, it was decided to use the Spearman rank order correlation as a preliminary test of the effect of the two scoring techniques. The

rejection of the first hypothesis would lead to an examination of how significant the relationships between sets of scores are.

The results of progress tests are used to predict achievement. Often this is merely an extension, by means of a fixed rank order in a class, to predict final examination standings, but sometimes simple or multiple regression equations are developed for specific institutions. The rejection of the second hypothesis would lead one to determine which of the techniques produced results which were most closely correlated with the results of the essay and problem-type mid-term examinations.

V. SUMMARY OF CHAPTER I

The purpose of this study was to collect and examine empirical evidence concerning a method of extracting additional information from the results of objective tests --both conventionally-scored standardized tests and those specifically prepared to exhibit the characteristics in question. The resulting data were correlated with regular examination results with a view to indicating the value of the application of this data to prediction purposes.

A further aim of the study was to indicate the application of the subjective probability scoring technique to diagnostic purposes in specific subject areas.

VI. OUTLINE OF THE PRESENTATION

The next chapter presents a review of the literature on the related fields of:

- (a) Objective test use;
- (b) Results of and corrections for guessing on objective tests;
- (c) The meaning of subjective probability;
- (d) The application of subjective probability scoring to indicate students' degrees of confidence by their weighting of answers to test items.

Chapter III outlines the design of the investigation and relates information on the population, instruments and procedure.

Chapter IV is an analysis of the data collected and includes both item analyses of the instruments and group data correlations.

The conclusions and implications of the study are presented in the final chapter.

CHAPTER II

A REVIEW OF THE LITERATURE

In order to assess the various testing techniques, particularly to indicate the subjective probability application, an examination was made of professional publications which included: The Mathematics Teacher, Journal of School Science and Mathematics, Journal of Experimental Education, Scientific American, British Journal of Psychology, Educational and Psychological Measurement, Journal of Educational Psychology, Psychometrika and Acta Psychologica. It was decided to make a four-staged approach to the topic:

1. An introduction of types of objective testing;
2. An examination of the question of guessing;
3. The meaning generally taken for "subjective probability;"
4. A review of what has been done and suggested with the subjective probability approach.

Much has been written and spoken concerning the relative merit of various types of test items. Adams (1954) has summarized the advantages of the various types of test items. Her conclusions indicate that for most purposes the objective tests are more appropriate than the essay or supply-type items.