

THE UNIVERSITY OF MANITOBA

THE REVISION OF AN EDUCATIONAL ACHIEVEMENT TEST
FOR THE FRENCH-SPEAKING SOLDIERS
OF THE CANADIAN ARMY

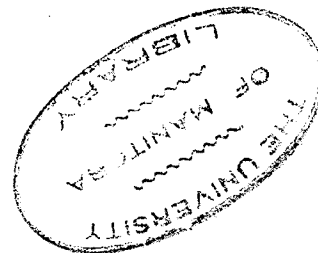
BEING A THESIS SUBMITTED TO THE FACULTY OF GRADUATE
STUDIES AND RESEARCH IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF EDUCATION

BY

LOUIS I. MASSON

MONTREAL, CANADA

DECEMBER 1956



ACKNOWLEDGMENTS

The writer wishes to express his feelings of deep indebtedness to:

Dr. H.L. Stein, his adviser, for his help and guidance.

Major A.B. Laver of Army Headquarters for his advice and help on statistical procedures.

Lt-Col. W.R.N. Blair and Major R.A. Wendt of Army Headquarters, for their help and encouragement.

Major J.A. Gagnon, Command Personnel Officer, Quebec Command, for his advice.

Col. H.M. Cathcart, Commander, Camp Valcartier, for his kind cooperation.

Mgr. R. Dolbec, Secretary General Laval University, and to the staff of "L'Ecole de Pedagogie et d'Orientation" for their cooperation.

The staff of the Psychology Department, McGill University, for its cooperation.

S/Sgt. Deneumoustier, Sgt. C.H. Morin, Miss E. Bradshaw, Miss G. Theoret and Mrs. M. Creasey for free time given to corrections, computation and typing.

Mr. Napier for the use of the "Marchant Calculator".

Mr. Crabb for the use of the "Monroe Calculator".

To all the Unit Commanding Officers of Camp Valcartier who provided the samples, and to all the NCO's and men who composed the samples and without whose cooperation this project could not have been undertaken and concluded.

TABLE OF CONTENTS

| | Page |
|---|------|
| List of Tables | iv |
| List of Figures | vii |
| Chapter | |
| I INTRODUCTION | 1 |
| Necessity of Educational Assessment | 1 |
| Early History | 2 |
| The Development of a Project | 2 |
| The Setting | 3 |
| The Method | 3 |
| II REVIEW OF ACHIEVEMENT AND SCHOLASTIC TESTING | 5 |
| III THE PROBLEM | 15 |
| History and Description of the Test | 15 |
| Shortcomings | 25 |
| Army Decisions | 26 |
| Administrative Arrangements | 26 |
| IV THE INVESTIGATION | 28 |
| Selection of Samples | 28 |
| Analysis of Samples | 29 |
| Administering and analysing the test | 35 |
| First Findings | |
| Comparisons with other findings | |
| Item Analysis | 40 |
| Method of Operation | |
| Item Difficulty | |
| Item Validity | |
| Rebuilding the Test | 46 |
| Second Administration | 48 |
| The Samples | |
| First Findings | |
| V THE RESULTS | 52 |
| Composition of the New Forms | 53 |
| Reliability of the New Forms | 75 |
| Problem of Validity | 81 |
| Establishing norms | 83 |
| Summary | 91 |
| VI CONCLUSIONS | 92 |
| The Achievement of a Purpose | 92 |
| The Advantages and Shortcomings | 98 |
| Problems for further Research | 100 |
| BIBLIOGRAPHY | 102 |

LIST OF TABLES

| Table | Page |
|--|------|
| 1 DISTRIBUTION OF TEST SCORES IN FIVE GROUP SCALE (ALL SERVICES) (1950 SAMPLE) | 17 |
| 2 DISTRIBUTION OF TOTAL SCORES BY GRADES IN FIVE GROUP SCALE (ALL SERVICES) (1950 SAMPLE). | 18 |
| 3 SUB-TEST AND TOTAL TEST RELIABILITIES WITH STANDARD ERROR OF MEASUREMENT (1949 SAMPLE). | 19 |
| 4 INTERCORRELATIONS BETWEEN FORMS "A" AND "B" | 20 |
| 5 RELATION OF TEST SCORES TO CLAIMED SCHOOL GRADES AND TO CLASSIFICATION TEST SCORES (1949 SAMPLE) | 21 |
| 6 DISTRIBUTION OF ITEM DIFFICULTIES BY SUB-TESTS FORM "A" (1950 SAMPLE) | 22 |
| 7 DISTRIBUTION OF ITEM DIFFICULTIES BY SUB-TESTS FORM "B" (1950 SAMPLE) | 22 |
| 8 DISTRIBUTION OF ITEM DISCRIMINATIVE VALUE BY SUB-TESTS FORM "A" (1950 SAMPLE) | 23 |
| 9 DISTRIBUTION OF ITEM DISCRIMINATIVE VALUE BY SUB/TESTS FORM "B" (1950 SAMPLE) | 23 |
| 10 COMPOSITION OF FIRST SAMPLES (BY RANK) | 29 |
| 11 COMPOSITION OF SAMPLES (BY CORPS) | 30 |
| 12 BASIC STATISTICS OF SAMPLES "A" AND "B" ON M SCORES, EDUCATION AND AGE | 31 |
| 13 COMPARISON OF SAMPLES "A" AND "B" WITH ONE ANOTHER AND OF INDIVIDUAL SAMPLES WITH LARGER GROUP OF RECENT ENROLEES (M SCORES) | 32 |
| 14 COMPARISON OF SAMPLES "A" AND "B" WITH ONE ANOTHER AND OF INDIVIDUAL SAMPLES WITH LARGER GROUP OF RECENT ENROLEES (EDUCATION) | 33 |

| | | |
|-------|---|----|
| 15 | COMPARISON OF SAMPLES "A" AND "B" WITH ONE ANOTHER ON AGE | 33 |
| 16 | MEANS AND STANDARD DEVIATIONS FORMS "A" AND "B" - TOTAL FORMS "A" AND "B" - SUB-TESTS | 35 |
| 17 | COMPARISON OF MEANS OF 1950 SAMPLE WITH PRESENT GROUP ON 2 SUB-TESTS | 36 |
| 18(a) | TOTAL - SUB-TEST AND INTER-SUB-TEST COEFFICIENTS OF CORRELATION (FORM A) | 38 |
| 18(b) | TOTAL - SUB-TEST AND INTER-SUB-TEST COEFFICIENTS OF CORRELATION (FORM B) | 38 |
| 19 | COEFFICIENTS OF CORRELATION BETWEEN EST AND M (BOTH FORMS) | 39 |
| 20(a) | DISTRIBUTION OF ITEM DIFFICULTY INDICES BY SUB-TESTS (FORM A) | 42 |
| 20(b) | DISTRIBUTION OF ITEM DIFFICULTY INDICES BY SUB-TESTS (FORM B) | 42 |
| 21 | DISTRIBUTION OF ITEMS BY SUB-TESTS (ORIGINAL AND REVISED FORMS) | 47 |
| 22 | COMPOSITION OF SECOND SAMPLES (BY RANK) | 48 |
| 23 | COMPOSITION OF SECOND SAMPLES (BY CORPS) | 48 |
| 24 | COMPARISON OF SECOND SAMPLE WITH FIRST SAMPLES AND WITH LARGER GROUP OF RECENT ENROLEES (M SCORES) | 49 |
| 25 | COMPARISON OF SECOND SAMPLES WITH FIRST SAMPLES AND WITH LARGER GROUP OF RECENT ENROLEES (EDUCATION) | 50 |
| 26 | MEANS AND STANDARD DEVIATIONS - FORMS "A" AND "B" (REVISED) - TOTAL - FORMS "A" AND "B" (REVISED) - SUB-TESTS | 52 |
| 27 | ITEMS ELIMINATED FROM VOCABULARY, FORM "B" WITH INTERNAL VALIDITY INDICES | 53 |
| 28 | REARRANGED ORDER OF ITEMS WITH DIFFICULTY VALUES VOCABULARY FORM "A" | 54 |
| 29 | REARRANGED ORDER OF ITEMS WITH DIFFICULTY VALUES VOCABULARY FORM "B" | 55 |

| | | |
|----|---|----|
| 30 | ITEMS ELIMINATED FROM READING COMPREHENSION SUB-TEST (FORMS A & B) WITH INTERNAL VALIDITY INDICES | 57 |
| 31 | REARRANGED ORDER OF SECTIONS WITH MEAN DIFFICULTY VALUES READING COMPREHENSION FORM A | 58 |
| 32 | REARRANGED ORDER OF SECTIONS WITH MEAN DIFFICULTY VALUES READING COMPREHENSION FORM B | 59 |
| 33 | REJECTED ITEMS WITH INTERNAL VALIDITY INDICES SPELLING (FORMS A AND B). | 60 |
| 34 | REARRANGED ORDER OF ITEMS WITH DIFFICULTY VALUES SPELLING FORM "A" | 62 |
| 35 | REARRANGED ORDER OF ITEMS WITH DIFFICULTY VALUES SPELLING FORM "B" | 63 |
| 36 | ITEMS ELIMINATED FROM SOCIAL SCIENCE SUB-TEST WITH FUNCTIONALITY INDICES (FORMS A AND B) | 64 |
| 37 | REARRANGED ORDER OF ITEMS WITH DIFFICULTY VALUES SOCIAL SCIENCE - FORM A | 66 |
| 38 | REARRANGED ORDER OF ITEMS WITH DIFFICULTY VALUES SOCIAL SCIENCE - FORM B | 67 |
| 39 | ITEMS ELIMINATED FROM GENERAL SCIENCE WITH INTERNAL VALIDITY INDICES FORM A & B | 69 |
| 40 | REARRANGED ORDER OF ITEMS WITH DIFFICULTY VALUES GENERAL SCIENCE - FORM A | 70 |
| 41 | REARRANGED ORDER OF ITEMS WITH DIFFICULTY VALUES GENERAL SCIENCE - FORM B | 71 |
| 42 | ITEMS ELIMINATED FROM MATHEMATICS WITH INTERNAL VALIDITY INDICES - FORMS A & B | 72 |
| 43 | REARRANGED ORDER OF ITEMS WITH DIFFICULTY VALUES MATHEMATICS - FORM A | 73 |
| 44 | REARRANGED ORDER OF ITEMS WITH DIFFICULTY VALUES MATHEMATICS - FORM B | 74 |
| 45 | RELIABILITY COEFFICIENTS - SUB-TEST AND TOTAL TEST - FORMS A AND B | 78 |

| | | |
|----|---|----|
| 46 | COEFFICIENTS OF CORRELATION BETWEEN EACH SUB-TEST AND THE WHOLE MINUS THE SUB-TEST FORM A . | 79 |
| 47 | COEFFICIENTS OF CORRELATION BETWEEN EACH SUB-TEST AND THE WHOLE MINUS THE SUB-TEST FORM B . | 79 |
| 48 | COEFFICIENTS OF CORRELATION BETWEEN REVISED EST, CLAIMED EDUCATION AND M TEST SCORES. . | 82 |
| 49 | DISTRIBUTION OF SCORES BY GRADES (CLAIMED EDUCATION) FORM "A" REVISED | 84 |
| 50 | DISTRIBUTION OF SCORES BY GRADES (CLAIMED EDUCATION) FORM "B" REVISED | 84 |
| 51 | PERCENTILE SCORES EQUIVALENT TO RAW SCORES - FORMS A & B (REVISED). . . . | 88 |

LIST OF FIGURES

| | |
|----------|----|
| Figure 1 | 89 |
| Figure 2 | 90 |

APPENDICES

A RAW SCORE DATA (Old Forms)
B RAW SCORE DATA (New Forms)
C STATISTICAL PROCEDURES
D THE "M" TEST
E THE EST (Forms "A" & "B")
F THE EST (Revised Forms)

CHAPTER I
INTRODUCTION

It has been a practice in the Canadian Army to take educational status into account in determining the suitability of individuals for enrolment. It has been a practice also to take educational status into account in selecting men for particular tasks. This practice was carried out during WW II but with the advent of Canada's peace time Army its necessity became more obvious.

However, the means by which educational status may be assessed are evidently numerous and varied. One can, for instance, assess educational status on the basis of former schooling records, scores on classification tests or on observation. Still, none of these methods would yield results which could be expressed in terms equivalent to job requirements. Moreover, school curricula vary from province to province while army requirements are national and circumstances of schooling, class standing, length of time since leaving school, and other factors probably have a bearing in each individual case.

It was with these ideas in mind that shortly after the end of hostilities (1947) the Defence Research Board decided to:

"prepare test materials which could be used in assessing the educational achievement of young Canadian adults and in revealing the individual's strong or weak position in this respect."¹

-
1. R.D. Dickie and B.K. Doane. Appraisal of Educational Achievement of Service Personnel. Operational Research Memorandum No. 20. Defence Research Board, Ottawa, 1952.

Following some experimental runs and after analysis and revision, the tests were evolved by 1950 and delivered to the services with tentative norms.

In 1950 also a contract was given to the University of Ottawa for the development of a French version of the English "Educational Survey Test".²

Work was commenced on the French version immediately, but this work came prematurely to a halt because of changes in responsibility for personnel research projects within the Defence Research Board.

As a result the present French version of the Educational Survey Test is made up of 2 sub-tests which were developed and standardized on a French-speaking school-children population, and of 4 sub-tests which are literal translations from the English version. When the changes of responsibility occurred, the University of Ottawa had completed its work only on 2 of the 6 sub-tests. Thus the French test which was delivered to the services was never standardized on a French-speaking adult population and none of its results was ever analyzed.

Naturally the question of continuing and future research on the French editions particularly was discussed several times.³ However, no systematic work was done in this area until October 1954 when Personnel Selection Service, at Army Headquarters, suggested to the writer that the present study should be undertaken.

2. R.D. Dickie. Personal correspondence with the writer. 1954.

3. Ibid.

The problem, evolved at a conference of the Personnel Selection Service Staff and of the writer in October 1954, was essentially two-fold:

First, an item analysis was to be made on the French version and all items in the test were to be closely checked in order that their respective difficulty and validity be determined, and, secondly, the test was to be revised in the light of the item analysis and re-administered to another group for the purpose of establishing norms.

This necessitated the administration of the two forms of the test to a group of French-speaking Canadian soldiers. It necessitated, also, the analysis of results and the development of the revised version, as well as its administration upon another French-speaking sample.

It was decided, then, that all samples should be taken from the population of the main French-speaking camp in Canada (i.e. Camp Valcartier) whose population, in the fall of 1954, was approximately 4000 all ranks. It was decided, also, that the sample should comprise about 4% of the total camp population and, that it should be representative of the French-speaking population of the Canadian Army as a whole at least on the factors intelligence and claimed education.

On the question of intelligence, records were available for any of the men to be selected through their results on the Army M test.⁴

On the question of education, again records were available through individual Personnel Selection documents.⁵ Thus the whole project included the following steps:

(a) A review of the history of educational assessment in the Army was to be made.

4. Revised Examination "M". Developed during WW II by the Canadian Psychological Association.
5. CAFB 1538. Personnel Selection Record.

- (b) Samples had to be selected for the administration of the original test and eventually for the revised one.
- (c) These samples were to be compared with the French-speaking population of the Canadian Army as a whole, on the question of intelligence and education.
- (d) The results of the first administration of the test were to be analysed.
- (e) The revised form of the test was to be evolved on the basis of the results obtained through the first administration.
- (f) The results of the second administration were to be analysed and norms were to be established on their basis.

The foregoing steps are described in this study in the following manner:

- (a) Chapter II gives a summary of the history of educational assessment in the Canadian Army.
- (b) Chapter III states the problem by giving the history and the description of the test, its short-comings and the administrative arrangement involved in this project.
- (c) Chapter IV describes the investigation itself by reviewing the selection of samples, the method of operation in the item analysis, in the rebuilding of the test, and in the evolving of norms.
- (d) Chapter V gives the result of the study, and
- (e) Chapter VI concludes with recommendations and with the statement of problems for further research.

CHAPTER II

REVIEW OF ACHIEVEMENT AND SCHOLASTIC TESTING

Mursell has defined the psychological test as:

"a pattern of stimuli selected and organized to elicit responses which will reveal certain psychological characteristics in the person who makes them."¹

Such a definition could almost apply as well to achievement tests, for, in effect, achievement tests as we know them today (i.e. standardized tests) are composed of stimuli, the responses to which reveal some characteristic of the subjects making them.

However, there is a broad distinction between the psychological test and the achievement test. The first measures certain psychological characteristics while the second measures knowledge in a certain area. Both, of course, involve mental processes, but, one is necessarily general in scope while the other is specific in that it measures specific areas of acquired knowledge. They also differ in their purpose, for the psychological test intends to measure one's ability while the achievement test measures one's accomplishment.

Still, it must be said that both converge towards one another and that they have much in common.

Consequently, as it is the purpose of this chapter to review educational measurement, it will be necessary at first to retrace some of the history of psychological testing in general.

1. James L. Mursell. Psychological Testing. Longmans Green and Co. New York 1949. p 1

The Origins of Psychological Testing

Psychological Tests originated with Alfred Binet, a French experimental psychologist, in the latter part of the last century. It was Alfred Binet who started the movement and his main contribution lay in the development of an intelligence scale which was to gain wide repute through the years that followed.

In the United States where the movement gained impetus rather rapidly, one of the most enthusiastic users of the Binet scales was Goddard whose revision of the Binet Tests appeared in 1911.² Kuhlman also started at an early stage to work with the Binet Scales and it was in 1922 that his "Handbook of Mental Tests" appeared.³ However, the main and the most well-known contribution came from Stanford University with Terman in 1916.⁴ There followed then, many tests devised for various purposes and among these emerged the Scholastic Achievement Test.

The Scholastic Achievement Test is that which is devised to measure educational knowledge, and, as it was mentioned in Chapter I, such was the purpose of the Army's Educational Survey Test.

In order to understand what is an "Educational Survey Test", one must go back and study the meaning of "Education" as well as that of "Measurement" within the framework of Education.

-
2. H.H. Goddard. A Revision of the Binet Scale. Training School Bulletin. 1911. Vol 8 pp 56-62.
 3. F. Kuhlman. A Handbook of Mental Tests. Warwick and York. Baltimore 1922.
 4. Lewis M. Terman. The Measurement of Intelligence. Houghton Mifflin Company. New York. 1916.

Education is defined in a broad sense by Travers as "all deliberate attempts to change behaviour".⁵ Such a definition of course encompasses all forms of education including that brought about by newspapers, radio, television, etc. Naturally, such a definition is not exactly satisfactory for the purpose of this study, although the education which is measured in the army has often been brought about through such media as those mentioned above.

More specifically though, Education may be defined for our purpose as "knowledge gained in various areas which are normally taught in Canadian schools."

Now, there are of course numerous means by which the extent of this knowledge may be appraised.

First of all, we may differentiate between "formal" and "informal" appraisal. Most of the appraisals that are made in schools today are informal (i.e. they are based on individual judgments of teachers, of supervisors or of other authoritative bodies).

The formal appraisal on the other hand is seldom met and for various reasons:

First, the cost of such appraisals is always substantial and the initial outlay of funds must be made often without assurance that the outcome will be satisfactory.

Secondly, educational measurement is a relatively new field and as yet measuring instruments are limited.

However, there is at present a trend in education towards more formal appraisal and assessment and as it was pointed out in Chapter I, such assessment is essential in a setting such as that of the army.

5. Robert M.W. Travers. Educational Measurement. The MacMillan Company. New York 1955. pp 181-182.

Now, through the years, many works have been written on the question of objective educational assessment and there have evolved certain guiding principles which we shall attempt to review in this study.

Of course, we are not concerned so much with the reaching of certain educational objectives, as one might be in a school, for the actual "educational" learning situations in the service are fewer. We are more concerned perhaps with the evaluation of the learning which has been effected in an environment that, for the soldier, preceded the military setting. Consequently, once again we are not concerned with changes occurring during service but rather with those that have occurred prior to service and which are reflected in an individual's knowledge.

Perhaps one of the outstanding landmarks in the field of educational measurement has been the "Pennsylvania Study".⁶ The Pennsylvania Study was conducted under the auspices of the Carnegie Foundation for the Advancement of Teaching by Wood and Learned. It had intended to review the system of schooling in Pennsylvania in the light of objective evidence concerning its obtained results. The authors had also hoped that a series of tests developed as a part of the study might become useful substitutes for the system of determining levels of educational achievement. Thus, in a way, these tests were intended to do the same thing as the Army's Educational Survey Test.

Essentially, the Pennsylvania Study appraised education in terms of knowledge, as does the Educational Survey Test, and the authors had defined their meaning of knowledge as follows:

"Thus, according to this view, all education is unavoidably intellectual. Its business is to make clear which ideas are true and valuable, which dubious or trivial, which deserve emotional support and why. But it is considered absurd for education to attempt to proceed without ideas, or to abjure the conviction that true and important ideas are the first requisite in guiding profitably the emotional life of a normal individual.

6. William S. Learned and B.D. Wood. The Student and his Knowledge. New York. The Carnegie Foundation for the Advancement of Teaching. 1938.

The inquiry which is now reporting subscribes to this same point of view. Fundamental to its procedure is the premise that education consists in thinking, in the perception of meaning and relationships among ideas which are true and important, and in the marshalling of an individual's natural emotions behind ideas in proportion to their truth and importance."⁷

Thus, one can see from this definition that the authors of the "Pennsylvania Study" had a very broad concept of knowledge, much broader in fact than that which the Army must have had when it developed the Educational Survey Test.

Still, the tests which were evolved did emphasize "information" and in spite of the author's definition of knowledge, there is no evidence in the content of the tests of an effort to measure depth of understanding. It must be admitted that some emphasis was placed on thinking skill, but that is as far as the tests went and one can hardly see how the "marshalling of an individual's emotions behind ideas in proportion to their truth and importance" is actually measured in the tests developed by Learned and Wood.

The results of the Pennsylvania Study, on the whole, were perhaps disappointing for the study had failed to show how objective tests could be used as a substitute of the system of awarding credit on the basis of number of hours spent in class.

Of course, it had been assumed that there is a relatively fixed body of knowledge with which everybody should have some familiarity but as the system of schooling differs from one place to another, so does the body of knowledge and, as a result, the assumption was untenable.

Such an assumption, of course, had also been made in the Educational Survey Test⁸ - although, recently, a directive pointed out its fallacy and gave a new definition to its purpose.⁹

7. Ibid. pp 5-6.

8. Personnel Officers' Guide. Section 6-10 (1952).

9. Personnel Officers' Circular Letter #4. 1955.

Perhaps the Pennsylvania Study has contributed more by its effects on American education than by its actual results. For one thing, it did give a tremendous impetus to educational testing in America and its authors extended their work to such widely known and very useful agencies as the Educational Testing Service which publishes the Cooperative Tests in various school subjects as well as to College level tests such as the "Cooperative
10
General Culture Test".

Now, in the Canadian Army, Education was never assessed formally until the Educational Survey Test appeared in 1950.

During World War II for instance, the Education of recruits or of serving soldiers was not measured at all, but simply recorded as given by the recruit or soldier himself, or from school reports if these were available. Naturally, with Officer Candidates, the appraisal of education was considered more important, but the literature does not reveal any attempt at measuring education in a formal way even with this group.

Still, today, it is common practice only to record education and to appraise it formally by the use of a test only in special cases.

Thus, the development of the Educational Survey Test was really the first attempt ever made to measure education in the army in a formal fashion.

In the United States Forces, a test had been devised at the end of World War II, the purpose of which was: "to help schools and colleges determine the amount of academic credit which should be granted students for their educational experiences in military service during World War II."
11

-
10. Cooperative General Culture Test. Princeton, New Jersey. Cooperative Test Division of the Educational Testing Service. Form X. 1944.
 11. Anastasi, Anne. Psychological Testing. The MacMillan Company. New York. 1955. p 471.

This test was called the United States Armed Forces Institute General Educational Development Test (US AFI GED).

The high school battery of this test consists of five tests covering the following areas:

- a) Correctness and effectiveness of expression
- b) General mathematical ability
- c) Interpretation of reading materials
- d) Social studies
- e) Natural sciences
- f) Literature

Norms were established on large samples of civilian students and for six different geographical sections of the United States.

This brief description of the test shows that it is very similar to the Army's Educational Survey Test - although the American test is much longer (each part requires 2 hours).

Although this test battery has been in wide use in the United States, it is agreed that a great deal of information is still needed on its empirical validity, on the intercorrelations of separate tests and on its correlation with other standard tests.