

DIFFERENTIAL RETENTION OF
COURSE OUTCOMES IN EDUCATIONAL PSYCHOLOGY

by
William P. McDougall

A DISSERTATION

Presented to the Faculty of
The University of Nebraska in the Teachers College
in Partial Fulfillment of Requirements
For the Degree of Doctor of Education
Department of Educational Psychology and Measurements

Under the Supervision of Professor Charles O. Heidt

Lincoln, Nebraska

1957

TITLE

DIFFERENTIAL RETENTION OF COURSE OUTCOMES

IN EDUCATIONAL PSYCHOLOGY

BY

William Phillips McDougall

APPROVED

DATE

Charles O. Neidt

July 26, 1957

Joseph L. French

July 26, 1957

Madison Brewer

July 26, 1957

Howard E. Tempero

July 26, 1957

M. F. Thorpe

July 26, 1957

SUPERVISORY COMMITTEE

ACKNOWLEDGMENTS

The author is very grateful to his major advisor, Professor Charles O. Heidt, Chairman, Department of Educational Psychology and Measurements, for his generous assistance and inspiration in the development and preparation of this thesis. Appreciation is also expressed to Dr. Robert T. Littrell who made the experiment possible by taking a major part of the responsibility for the collection and processing of the test data.

The author also wishes to thank the instructors in the Department of Educational Psychology and Measurements, Elementary and Secondary Education, for their cooperation by allowing class time for the collection of the data.

W.F.H.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	i
TABLE OF CONTENTS	ii
LIST OF TABLES	iv
LIST OF FIGURES	v
 I INTRODUCTION	 1
A. Statement of Problem	1
B. Review of Related Literature	3
1. Instructional Objectives	3
2. Retention Studies	10
C. Need of Study	15
 II PLAN AND PROCEDURE	 17
A. Trial Test Construction	17
1. How the Trial Test was Constructed	18
2. Description of the Test Items	19
a. Knowledge	20
b. Comprehension	21
Translation	21
Interpretation	22
Extrapolation	23
B. The Trial Test	24
1. The Trial Test Group	24
2. Trial Test Administration	24
C. Refinement of the Test	25
1. Item Analysis	25
2. Evidence of Reliability	26
3. Homogeneity of Test Behavior	27
D. Administration of the Refined Test	29
1. Description of the Test	29
2. Description of the Subjects	30
E. Treatment of the Data	31
1. Item Difficulty and Discrimination	31
2. Reliability of the Tests	31
3. Homogeneity of Tests	32
4. Evidence of Validity	32
5. Retention	33
 III FINDINGS OF THE STUDY	 37
A. Analysis of the Tests	37
1. Item Difficulty	37
2. Item Discrimination	38
3. Evidence of Reliability	41
4. Homogeneity of Tests	42
5. Evidence of Validity	46

B. The Study of Retention	47
1. The Retention Group	48
2. Relationships Between Pretest - Test - Retest	49
3. Differences in Average Performance for the Various Test Administrations	50
4. The Retention Graph	51
5. Average Percent of Gain Retained	52
6. Occurrence of Certain Patterns on the Successive Administrations of the Test Items.	54
IV IMPLICATIONS OF THE INVESTIGATION	56
V SUMMARY	59
SELECTED REFERENCES	63
APPENDIX A	66
APPENDIX B	85
APPENDIX C	88
APPENDIX D	100
APPENDIX E	101

LIST OF TABLES

	Page
Table 1 Means, Standard Deviations and Spearman-Brown Estimates of Reliability for Tests	27
Table 2 Intra- and Inter-Correlations Between Tests	28
Table 3 Values of F Between Tests for Departure from Homogeneity	29
Table 4 Item Difficulty and Item Discrimination for Test Score and Combined Total Score	39
Table 5 Spearman-Brown and Kuder-Richardson Estimates of Reliability	42
Table 6 Intra- and Inter-Area Correlations Between Tests . .	44
Table 7 Values of F for Tests of Homogeneity Between Tests .	44
Table 8 Values of F for Tests of Homogeneity Between Tests (Interpretation-Extrapolation Combined)	45
Table 9 Correlation Coefficients Between the Three Administrations of the Tests	50
Table 10 Differences in Mean Scores and the Accompanying t-value Between the Three Administrations of the Tests	51

LIST OF FIGURES

	Page
Figure I Retention Graph of the Three Tests: Knowledge, Translation, Interpretation-Extrapolation	52

PREVIEW

CHAPTER I

INTRODUCTIONA. Statement of Problem

Tyler stated: "...the organization of courses and the development of examinations should center about those objectives having more permanent values."^(35:203)¹ Of major concern to the classroom teacher is the need for accurate evidence concerning the degree of permanency of each instructional objective. Such information would certainly enable the teacher to function more effectively in guiding the learning activities of students.

The relative permanency of different types of learning has often been overlooked. Frequently it has been assumed that the results of the usual instructional procedures are of lasting value and that the attainment of one objective can be inferred from measured attainment of another objective. The practices of defining each educational goal, validating evaluation devices against defined aims, and studying the attainment of each one of these aims to determine its degree of permanency are clearly needs in our instructional programs.

The purpose of this study was to measure retention of different course outcomes in a beginning course in educational psychology. The outcomes examined included knowledge and the intellectual skills and abilities specified as translation, interpretation and extrapolation. These objectives were defined by the "Taxonomy of Educational Objectives"⁽⁶⁾,

1. The number which appears first in the parenthesis refers to a title in the list of references. The number following the colon indicates the pages on which the citation appears.

and are explained briefly in the following paragraph. The subject matter content studied was delimited to one area, tests and measurements, to permit more intensive and uniform sampling of the different course outcomes. The samples of responses to the test situations were taken from the Education 62 classes at the University of Nebraska. (Education 62, Human Behavior and Development, is the second of two courses in educational psychology taken by teacher trainees.) Follow up testing was done approximately a semester later in the teacher training sequence.

According to the Taxonomy of Objectives, six major hierarchical levels of educational objectives are defined in the cognitive domain. Of these six the first two, specified as knowledge and comprehension probably include the largest general classes of intellectual activities emphasized in schools and colleges. (6:89) For this reason these two major categories were selected for this study. The category of knowledge includes recognition and recall of specifics, ways and means of dealing with specifics, and knowledge of universals and abstractions. Test items designed to measure these categories seem to be very similar and, therefore, all knowledge items were defined as being in a single category. The comprehension level includes items designed to measure the ability to translate knowledge from one form to another, to interpret data and to extrapolate from data. Items designed to measure the sub-categories on the comprehension level were grouped as separate tests. A more complete description of these behavioral objectives as defined in the Taxonomy is given in Chapter II in connection with the trial test construction.

The first task undertaken in this study was the construction and validation of the tests to be used to measure the desired outcomes. These tests were constructed and analyzed both in trial and final test form

using the following criteria:

1. The relationship of each test item to explicitly defined objectives.
2. The contribution of individual items to the effectiveness of its respective test.
3. The degree to which the separate tests measure the desired behavioral functions.
4. The degree to which the instrument satisfies conditions of empirical validity and reliability.

The second major part of the study consisted of the use of the tests to determine the degree to which the separately evaluated abilities were retained. For this purpose the tests were administered at three times: as a pretest at the beginning of the course, a post test as part of the final examination and again as a retention test approximately four months later. The differences between the sets of scores were then examined to ascertain the degree of permanency of each of these course outcomes.

3. Review of Related Research

1. Instructional Objectives

Examination of the literature clearly points out the need for careful planning and formulation of objectives in the construction of achievement tests. A summary of steps suggested by Tyler emphasizes this point. The steps are as follows: (24:5)

- 1) Formulation of course objectives
- 2) Definition of each objective in terms of student behavior
- 3) Collection of situations in which students will reveal presence or absence of each behavior

- 4) Presentation of situations to the students
- 5) Evaluation of students' reaction in light of each objective
- 6) Determination of objectivity of evaluation
- 7) Improvement of objectivity when necessary
- 8) Determination of reliability
- 9) Improvement of reliability when necessary
- 10) Development of more practical methods of measurement when necessary

The importance of objectives has been effectively stressed in the writing of Peters (26:148) who showed that few tests were validated against educational objectives and suggested that they should be derived on this basis. Jordan also emphasized this point of view (20:5), and stressed the necessity that the objectives of education be clearly defined or else the appropriate measuring instruments cannot be constructed. Other writers such as Bean (4:17), and Vaughn (22:160) have stressed the basic importance of delineated objectives.

Vaughn stated that too often a test outline is confined to subject matter alone and the type of behavior to be examined is left to the judgment of an inexperienced item writer. He indicated that a test outline can be used to show clearly not only what different areas of subject matter are to be covered but also the types of behavior to be elicited with respect to each area. Rogers (23:47) concurred with this point of emphasis and stated that the distinction between general and specific objectives should be made to facilitate the comprehension of objectives. He further pointed out that objectives should be stated in terms of observable changes in pupil behavior. Likewise, Tyler (23:30) emphasized that the most useful form for stating objectives is to express them in

terms which identify both the kind of behavior to be developed in the student and the content area of life in which this behavior is to operate. He stated that clearly formulated objectives included both the behavioral aspect and the content aspect. Tyler pointed out that objectives are often stated in terms of things which the instructor is to do, lists of topics, concepts, generalizations, or other elements of content to be dealt with, or as generalized patterns of behavior which failed to indicate more specifically the area of life or the content to which the behavior applied. He further stated that since the real purpose of education is to bring about changes in the students' pattern of behavior, it becomes clear that the statement of objectives should be in terms of behavioral change.

Flanagan (11) proposed that a "method of rationales" be employed for the purpose of clear and precise definitions of what is to be measured. The method begins with a list of behaviors to be sampled or predicted. The development of these rationales consists of three parts.

- 1) Description of behavior
- 2) Analysis of behavior
- 3) Formulation of item specifications

Description of behavior involves the definition, delimitation and illustration of the variety and scope of the actions included. Analysis of behavior includes clarifying it with respect to other behaviors and making inferences about its nature, culminating in the formulation of one or more hypotheses regarding its generality and predictability. Formulation of item specification carries the procedure on to describing a specific type of item which, it appears, should provide a valid estimation of the specified behavior.

Micheels (25:92) also emphasized the importance of the behavioral aspect of objectives. He proposed four steps to follow in setting forth the test objectives. The steps are:

- 1) List the major objectives for which appraisal is desired.
- 2) Examine the course content for additional objectives.
- 3) Analyze and define each objective in terms of expected student outcomes.
- 4) Establish a table of specifications for the tests.

He suggested that step three is the inventory step. Various elements are listed that are a part of each objective and meaning is given to each element by defining it in terms of student behavior.

Much attention has been given in the literature to the listing and categorization of the broad general outcomes of the educative process.

(2:14) Most frequent among the outcomes listed are such types as skills, knowledges, concepts, understandings, applications, attitudes, interests and adjustments.(16:168) (14:16)

In contrast to plans involving broad general classifications, other plans have been proposed which require more specific definition of objectives. The report, "A Design for General Education"(2:31), illustrated how the objectives in various fields can be classified into three basic categories: 1) knowledge and understanding, 2) skills and abilities, and 3) attitudes and appreciations. The authors of this report asserted that the three basic categories could be used as subdivisions for more general educational outcomes. The objectives for general education are organized under ten broad outcomes and in turn these are defined more specifically under the above listed categories. Approximately two hundred sub-objectives presented are useful in defining more specifically subject matter content and behaviors involved in general education.

Tyler illustrated the use of a two dimensional chart in stating objectives for a course (33:32) in biological science. One dimension of the chart defines the content aspect of the objectives, the other the behavioral. There are seven types of behaviors specified in the biological science course used as an illustration. Tyler describes these as follows:

"The first type of behavior is to develop understanding of important facts and principles. The second type is to develop familiarity with dependable sources of information... The third type of behavior is to develop ability to interpret data - that is, to draw reasonable generalizations from the kinds of scientific data likely to arise in this field. The fourth type of behavior is to develop ability to apply principles that are taught in biological science to concrete biological problems that arise in everyday life - hence to be able to carry on problem-solving activities in this field. The fifth type of behavior is to develop the ability to study and report the results of study. The sixth is to develop broad and mature interests as they relate to biological science, and the seventh is to develop social rather than selfish attitudes in this area."

The chart also includes a statement of the content aspects of the activities. Examples of these are such headings as nutrition of human organisms, digestion, circulation, respiration and reproduction. The course is then viewed as developing various sorts of behavior in relation to these aspects of content.

The "Taxonomy of Educational Objectives" includes the most careful definitions of objectives yet made. (6) Over several years a group of college and university examiners attacked the problem of creating a comprehensive classification of educational objectives. (6:4) Several uses are proposed for such a classification. The device is intended to facilitate communication among teachers and examiners, provide a framework upon which to plan learning experiences and prepare evaluation devices as well as to serve other useful educational needs. The efforts of this project resulted in the publication of a Handbook #1, which consists of a detailed

logical and psychological classification of educational aims in the cognitive domain. As the classification is now organized it contains six major classes: (6:16)

- 1.00 Knowledge
- 2.00 Comprehension
- 3.00 Application
- 4.00 Analysis
- 5.00 Synthesis
- 6.00 Evaluation

The authors conceive of these classes as representing a general hierarchical order. The behaviors of one class are likely to make use of and be built on behaviors found in the preceding classes. Since extensive use was made of the Taxonomy in the present study it will be described in greater detail in Chapter II.

Whereas the foregoing literature pertains to the general process of establishing objectives, authors of other publications stress the importance of clearly specifying the type of objectives being measured. Tyler (21:6) has studied the relationships between the recall of information on tests, the application of principles and the ability to draw inferences. These three types of tests were administered to sixteen different college classes in eleven different subject matter areas. The average number of students in the classes was 217 and ranged from twenty-two to 624. The resulting correlations between recall and application items ranged from .31 to .50 and were centered about .45. Correlations between the recall and inference items ranged from .27 to .60, averaging about .35. The ability to apply principles and to draw inferences yielded an average correlation of .40, ranging from .33 to .54. Tyler concluded that students did not

develop corresponding degrees of achievement in mere recall and in achievement in such higher mental processes such as the application of principles and the ability to draw inferences.

The results of McConnell's study (24:78) at the University of Minnesota agree with the work done by Tyler. McConnell dealt with tests in three subject matter divisions and achievement of three kinds of objectives; namely, knowledge of vocabulary, knowledge of facts and principles and ability to apply facts and principles. His analysis by the method of intercorrelation yielded an average correlation of .66. He also studied the extent of differential measurement by the method of item discrimination.

The assumption underlying this procedure was that an item designed to discriminate between students of high and low ability on one test would do so best when its own test total was used as the criterion rather than the total of a test designed to measure another objective. He found, using this method, that the average of all correlations within the three subject matter areas was .67 and the average among the three divisions was .66. This difference is statistically significant and means that there is relatively greater homogeneity of behavior within objectives than among objective groupings.

Other similar studies have resulted in comparable findings. Bodell (5) using a sample of 324 pupils in ninth grade general science found a correlation of .65 between performance on items requiring students to recall and to infer from science principles. Johnson (19) reported slightly higher coefficients, averaging about .55, between knowledge of fact and principles and the ability to apply them to new situations in college science courses. Brown (25:50), in the area of home economics, reported

that the correlations between a knowledge of scientific principles of cookery and the quality of food cooked or ability of people to manage their work was less than .50. Horrocks (16) studied the relationship of knowledge of facts and principles and the ability to apply them in the area of adolescent development. He utilized a criterion test of knowledge and three case study tests of application in testing three hundred upper classmen and graduate students. Each of the case study tests dealt with a different problem. The problems were in the social, academic and emotional areas. Resulting correlations between the criterion test and each of the case studies were .46, .41 and .26, respectively.

From the foregoing studies it is apparent that achievement of one objective cannot be inferred from the achievement of another. Remmers has expressed this point when he concluded: (28:31)

"...the educator must clearly define each objective in terms of the measure of its attainment. The attainment of a particular objective cannot be inferred from measured attainment of another objective."

2. Retention Studies

Though a large number of retention studies for many different school subjects have been reported, few have dealt directly with the problem of measuring the relative permanency of different kinds of course outcomes. In general, the studies show that forgetting proceeds rapidly at first and then more slowly. Much of what is shown by the usual final examination at the end of a college course is forgotten within four to six months. (27:544) Permanency of some outcomes seems to be, however, much greater than others. The first few studies will indicate the general nature of retention in college courses. The latter studies will treat the problem of differential permanency of instructional objectives.

A study of the retention of information learned in college courses was reported by Greene.(15:262) Retention of information demonstrated on the final examinations was studied in zoology, psychology, and physiological chemistry at intervals ranging from four to twenty months. The content of the final examination placed emphasis upon the recognition and recall of specific information rather than problem solving or logical organization of material. The final examination, which was given in June, was readministered to thirteen students in zoology, twenty-six in psychology and eighty-eight in chemistry during the following October. Allowing for initial learning, approximately one-half of the material reported correctly on the June final was lost in the four-month period. The proportion of the final examination scores which was retained eight months later averaged roughly one-quarter for zoology and one-fifth for psychology, and at the end of twenty months this averaged from one-tenth to one-fifth. Chemistry was not studied beyond the four-month period. Greene also reported correlations between the June and October scores for fifteen subjects in zoology, twenty-six in psychology and eighty-eight in chemistry. The resulting correlations were .706, .412 and .434, respectively. He suggests that the higher correlation in zoology may be attributed to the fact that it was a laboratory course and may have resulted in more consistent retention than a lecture course.

Watson (40) studied the permanence of material learned in a required introductory psychology course. One hundred college students were tested by typical examination questions for immediate recognition and recall and then, having been divided into six highly comparable subgroups ranging in size from ten to thirty-three, were tested for delayed retention after intervals of two, four, six, eight, ten, eighteen, twenty, twenty-two,

thirty, thirty-two, thirty-four, forty-two, forty-four, forty-six, fifty-four, fifty-six and fifty-eight months. Each subgroup was tested for three delay periods, a test of different type being used for each interval. Watson concluded that although forgetting increased with time, the point of complete forgetting was not reached even after fifty-eight months. Different results were obtained for recognition and recall type items. After a delay of two and one-half months he reported that ninety-seven percent retention occurred on the recognition scores; after a period of twelve months eighty-four percent occurred. The material learned that required recall in the test situations yielded much lower percentages. For the two and one-half month delay the percentage was thirty-seven, for six months thirty-four, and for nine months thirty-three percent. In general, the recognition curves decreased gradually and progressively and the recall curves decreased abruptly and progressively throughout the delay period.

A study of retention of knowledge acquired in a course in general psychology reported by Lurich (9) showed a higher degree of permanence of learning. Two final examinations constructed by the Department of Psychology at the University of Minnesota were used separately to study retention over periods of six and nine months. The one examination consisting of single choice, analogy, wrong word and completion type items was given at the completion of a psychology course and again nine months later to ninety-nine students. The other similar test contained no completion items but did include matching items and was given as a retest to eighty-three individuals six months later. The tests apparently measured largely factual information but this cannot be determined from the report. The mean score nine months after course completion was approximately

seventy-five percent of the mean at the close of the course. On the other retest after the six-month period, the mean score was ninety percent of the mean at the close of the course. Murich concluded that retests in psychology show that students retain a substantial amount of the measured knowledge six and nine months after they have taken the course.

The following studies will review findings where attempts have been made to identify separately the permanency of different types of course outcomes.

Tyler (35) studied the scores of eighty-two typical students in a course in elementary zoology who had been given several types of examination exercises at the close of the course and again fifteen months later. Five types of performance were measured. These were: naming animal structures pictured; identifying technical terms; recalling factual information; applying principles to new situations; and interpreting new experiments. Seventy-seven percent of the gains made during the course in the ability to name animal structures was lost in the fifteen-month period following. For the abilities to identify technical terms and recall information, approximately one-fourth of the gain made was lost. The ability to apply zoological principles to new situations or to interpret data from experiments showed no loss over the fifteen-month retention period even though these students had taken no other course in zoology during this time.

An experiment by Hart (41), also in zoology, yielded similar results. Separately measured course objectives were studied over a period of three years. The ability to apply principles and to interpret new experiments demonstrated in three years a percentage increase, fifty-eight and nineteen percent, respectively, over the students' performance at the end of

the course. There was a loss in the gain made during the course of over fifty percent in ability to remember terminology, functions of structures and main ideas. Over eighty percent loss occurred in associating names with structures. Wertz's study showed that retention was greatest in areas involving application and interpretation and least in areas involving informational objectives.

Five different objectives in high school chemistry were chosen by Frutchey (12) to be studied. These objectives were: selection of facts; application of principles; terminology; symbols; formulas, valence and balancing equations. Tests measuring these objectives were given to an average of eighty-three students as a pretest, again nine months later at the end of the course and a third time one year after completion of the course. The retention results were reported in terms of the percent of gain made in the course that was retained. For selection of facts eighty-four percent was retained, ninety-two percent for application of principles, sixty-six percent for terminology and about seventy-one percent for symbols, formulas, valence and balancing equations. Frutchey concluded that retention was greatest in the more general types of behavior.

Differential retention was also reported by Tyler in ninth grade general science.(36) Sixty-eight pupils were given the Ruch-Popence test, largely a measure of information, a multiple-choice explanation test, and a test requiring students to generalize from given facts. The tests were repeated eight months after the completion of the course. The mean loss for the Ruch-Popence was 11.1, for the explanations test 1.8, and for the test requiring generalizations 1.2. Apparently the ability to explain and generalize from known facts was retained better than mere information.

Individual differences in retention of general science subject matter involving recall of factual information, ability to explain scientific phenomena and ability to draw conclusions from given data were reported by Ward.(39) The tests measuring each of these behavioral outcomes were administered in June and again in September, approximately three and one-half months later to sixty-three students. The percentage of forgetting during this period was greatest for the factual information and conclusion parts, being approximately seventeen percent relative loss. For the test requiring the students to explain scientific phenomena, the retention was greatest with a mean loss of 9.1 percent. Ward felt that the findings of this study strongly suggest that ability to apply principles and to explain phenomena, and problem solving procedures, are retained over a long period with slight loss. He concluded also that the results reinforce a commonly accepted belief among educators that the permanent outcomes of teaching are to be found among the so-called "intangible objectives."

C. Need of Study

The need of the present study is suggested by the following quotation from the Taxonomy.(6:23)

"For the most part, research on problems in retention, growth and transfer has not been very specific with respect to the particular behavior involved. Thus, we are usually not able to determine from this research whether one kind of behavior is retained for a longer period of time than another or which kinds of educative experiences are most efficient in producing a particular kind of behavior. Many claims have been made for different educational procedures, particularly in relation to permanence of learning; but seldom have these been buttressed by research findings."

Judd has said,(21:4) "If by any means the educational system can discover how to promote even in the slightest measure the development of

the higher mental processes, great advantage will be gained for civilization." A first step in giving prominence to the so-called higher processes is to identify clearly which of these types of learning are of most lasting value. The review of literature in the previous section certainly strongly suggests such abilities as making inferences and applications stand out as demonstrating greater degrees of permanence than more factual information. The need to study other instructional outcomes such as those being studied in this investigation then becomes increasingly evident. In addition, the Taxonomy becomes a highly useful instrument to standardize the description of research findings and to facilitate communication of these results. Such behaviors as are studied can be much more clearly and universally identified, a task that has heretofore been difficult.

CHAPTER II

PLAN AND PROCEDURE

The general plan of this study involved the construction of tests to measure a variety of educational objectives in the content area of test and measurements. These tests were then used to study the relative retention of different abilities. More specifically, items were designed to measure each of the following:

- 1) the ability to recognize or recall basic knowledge
- 2) the ability to translate this knowledge from one form to another
- 3) the ability to interpret data
- 4) the ability to extrapolate from data

These tests were then administered as a trial test to a group of educational psychology students who had completed units on tests and measurements. The tests were then refined and used as instruments to study the retention of the above mentioned abilities. The refined test was given as a pretest¹, a test at the completion of the course and a re-test approximately a semester later.

A. Trial Test Construction

The trial test contained approximately thirty multiple-choice items in each of four categories - knowledge, translation, interpretation and extrapolation.

1. For purposes of this study, "pretest" will be used to describe the administration of the test at the beginning of the unit, "test" the administration at the end of the unit and "retest" the administration at the end of the four-month retention period.