SOME EFFECTS OF EMPHASIZING THE LEARNING FUNCTION OF CLASSROOM ACHIEVEMENT EXAMINATIONS

Thesis for the Degree of D. A. G.

MICHIGAN STATE UNIVERSITY
ABEL EKPO-UFOT

1969

ABSTRACT

SOME EFFECTS OF EMPHASIZING THE LEARNING FUNCTION OF CLASSROOM ACHIEVEMENT EXAMINATIONS

By

Abel Ekpo-Ufot

How may achievement examinations be conducted so as to better define and attain the objectives of classroom instruction? That is the problem investigated in this study. It is suggested it may be solved by emphasizing the learning function of examinations.

The suggestion rests on the literature evidence that examinations are a learning device. The authors quoted as supporting this view include Jersild, Standlee, Fitch and Page. This study was designed to test methods which might capitalize on such learning function.

The methods consisted in requiring one experimental group to take examinations twice--in class and outside class. A second experimental group both repeated and evaluated their performance before they had feedback. A third group, the control, was permitted to keep the test scripts. The hypotheses were that each of the experimental groups would score higher than the control on the final examinations, and the

second experimental would score significantly highest among the three. Furthermore the attitudes of the experimental groups would be more favorable towards examinations and grading than those of the control.

About 1300 students formed the population for the study. They were enrolled in two courses in the College of Education, Michigan State University during fall term 1968. In one course there were thirty-four classes, of whom thirty-two were randomly selected to run experiment 1. All the twelve classes in the other course were used in experiment 2. In both cases the classes were randomly allocated to treatment.

The study began with the development of a scale for measuring students' attitudes towards examinations and grading. Free response opinions were obtained from a sample of students by the use of an open-ended questionnaire. Content analysis of the returns produced nuclei statements for the attitude items.

Attitude is multidimensional. The key dimensions are "positive" and "negative," but these may not be on the same continuum since they are supported by different attributes of the psychological object. Such is the framework which guided the writing of attitude statements. These were tried out with a sample of 585 students representative of the university; the responses were factor analyzed and the final items selected on the basis of their loadings.

In the main study two examinations were administered within the term and students carried out instructions as

specified for their treatment conditions. The final examinations were the criterion measures of achievement. The Friedman X_r^2 test showed an overall treatment effect only in experiment 1, and a t-test revealed that the mean for the second experimental group was significantly larger than the one for the control. However, the absolute differences were small; but the trend was in the predicted direction. This is explained as due largely to the effect of the second treatment condition: it did stimulate effort and the self-evaluation would aid understanding.

The chief weakness of the study was poor control: all groups kept the within-term examination scripts. Also, the period of one term might have been too short for the treatment to work. These might partly account for the haphazard results obtained on the attitude criterion measures.

The main conclusion is based on the trend in the predicted direction revealed in experiment 1. If students were required to repeat and evaluate their examination performance their achievement of course objectives would tend to rise higher than what it would be without such conditions. It would be worthwhile to investigate the hypothesis that this "self-evaluation" would make students attitudes more "positive" than "negative" towards examinations and grading. Any contributions of this study to knowledge are conditional on collaborating evidence--evidence to support the usefulness of the attitude scale, and the model on which it is based, and

above all evidence to show that emphasis on the learning function of examinations will produce the type of effect weakly indicated in experiment 1 of the present study.

SOME EFFECTS OF EMPHASIZING THE LEARNING FUNCTION OF CLASSROOM ACHIEVEMENT EXAMINATIONS

By

Abel Ekpo-Ufot

A THESIS

Submitted to
Michigan State University
reporting research done as part
of the special requirements
for the degree of

EDUCATION SPECIALIST

College of Education

1969

This Thesis is Dedicated

to: Ufot-Ekpe--my late father,
Amma-mmi--my mother,

Fred Akpan--my uncle and stepfather,

Udo-Eka-Ekpo, Udo-Aka ("Gabriel"),
Ebenge and Iboro--my brothers,
Idorienyin--my sister,

to: Esit-Ima (Grace)--my wife,

and to: James S. Karslake--my strategist.

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued Page

LIST OF TABLES

LIST OF FIGURES

PREFACE

One conducting this type of study must have a bias; so has the writer. He does not share the view that classroom achievement testing should be abolished at a formal institution of learning--be it a school or a university. He does not consider that these twin aspects of the curriculum are a necessary evil: it may not be necessary that they should be to the student a "traumatic experience". Rather he shares the view that they are "a natural part of the total learning process."

But this bias may have intruded itself unwittingly into the tone of the presentation of this thesis; for this the writer sincerely apologizes to the reader. He really meant to present it as a scientific study uncolored by his biases--but he may not have succeeded. In particular he has offered tentative conclusions based on trends revealed in one of the two experiments conducted. But he has not hidden the fact that the evidence is very weak, not only because the absolute differences in the so-called trend were very small despite the "significance" of the t-test comparisons, but also because the results in the second experiment were not definitive. The reader should therefore take as

hypotheses to be investigated all tentative conclusions made in this thesis.

Many people have contributed to make this study possible. The twenty-two professors who undertook to administer the Attitude Scale deserve first mention; so also the students who served as "guinea pigs". The writer wishes to express his thanks to all these unnamed persons.

Thanks are due to Dr. Andrew C. Porter, and his staff of research consultants in the College of Education. Their criticisms and comments on the design of this study were of great value. Two of the writer's teachers deserve special mention: one is Dr. Maryellen McSweeny, of the College of Education, and the other Dr. Charles F. Wrigley, of the Psychology Department and Director of the Computer Institute for Social Science Research. As the reader will soon find this study was in a way a "try-out" of some research methods. These two professors gave the writer, among other students, a brilliant introduction to these methods. Besides they have criticized parts of the study that relate to their specialties, and in some cases have actually helped in the interpretation of the data. Another professor who had criticized parts of this study is Dr. Willard G. Warrington, Director of the Evaluation Services. His searching questionings contributed much to the development of the Attitude Scale to be reported.

The four members of the Program Committee occupy a unique position. The Chairman Dr. Robert L. Ebel, has indirectly inspired this study in that his philosophy is behind it. Dr. James S. Karslake, of the Psychology Department, urged that a research study be included in the writer's program for the Education Specialist degree. The other members are Drs. Robert C. Craig, Chairman of the Department of Educational Psychology and Dr. Paul L. Dressel, Director of Institutional Research. These four have each rendered constructive and valuable criticism on the thesis to be presented. The writer is grateful to them for their services.

Without a scholarship grant by the home Government of Nigeria the writer might not have embarked upon graduate education. This Government has therefore contributed indirectly but significantly to this study.

Thanks are also due to Drs. W. Sweetland, and D. Freeman, and their staffs of instructors and secretaries. The study might have been sabotaged without the cooperation of these people running the courses in which the experiments were carried out.

Apart from Dr. Karslake the other persons to whom this thesis is dedicated are of the household in which the writer is a member. He is deeply grateful to them for the price they are paying--to wait.

The last offer of thanks is to those who at one time or another have grappled with the problems of education. There is nothing reported here which is not owed to MAN.

CHAPTER I

INTRODUCTION

The focus of this study is on classroom achievement examinations and the twin practice of grading. In this introductory chapter some functions of examinations are stated. The fact that these may not always be realized leads to a statement of the problem to be investigated. This in turn is related to the evidence in the literature, in particular that which supports the view that examinations perform some learning functions. The chapter closes with a statement of the purposes and hypotheses of the study.

The Function of Examinations

An important objective of formal education is the acquisition of knowledge. Though, practices differ, classroom achievement examinations are widely used for assessing how far this objective has been attained. Examinations perform other important functions also. They motivate the student to learn. Admittedly, this function is differential; as Tyler (1966) observes they may stir up "feelings of incompetence in some students." However, it is likely that such unmotivated

students are in a small minority. Furthermore, examinations provide a learning experience *per se*: in Stone's (1955) words they "represent practice sessions which aid the fixation of correct responses . . . and the elimination of error." In other words, the taking of examinations in effect promotes and guides learning.

Moreover, in a society such as ours, it would appear one cannot escape evaluation. If the school exists to prepare youth to fit a need in society then it cannot altogether ignore some preliminary evaluation of the products it turns out to society. It may even be argued that such evaluation helps to remind the student of his future role, and that society expects him to be proficient in his fulfilling that role. If this argument be granted then, from the student's point of view, there are at least four functions of classroom examinations. The motivation for learning, the promotion, fostering and guiding of learning, the assessment of what "amount" has been learned and the reminder that learning must be proficient if one is to fulfill his role adequately in society--all these are of special importance to the student.

Dissatisfaction With Testing

Teachers tend to overemphasize the assessment function at the expense of other functions of classroom achievement testing. In such a situation, the attainment of the educational objectives may be limited or thwarted. Evidence is not

lacking that there is some dissatisfaction with testing in general and achievement testing in particular. Take for an example Banesh Hoffmann's book: The Tyranny of Testing (1962). The author is directing his attack against the "professional testers" and their reliance on multiple-choice tests and item statistics. But his view that "there is no satisfactory method of testing" applies to the classroom situation as well. "If sample questions made by the best test makers can give cause for concern," he asks, "What of multiple-choice tests made by individual teachers for their own classroom use. . .?"

The poor quality of test items, as Hoffmann says is cause for concern. But one is tempted to express the opinion that, within the classroom, the "tyranny of testing" is most evident in the teacher's too-much-emphasis on the assessment function of examinations at the expense of the learning functions, and the dissatisfaction among students may partly be explained by this fact.

The Problem

Granted that examinations serve important functions to the learner, it would appear there is a strong case to retain the system as an aspect of the school curriculum. If one takes this position he is faced with a problem: how may achievement examinations be improved in use so as to better define and attain the objectives of classroom instruction? This appears to be an important practical problem in all

education. It may be that the achievement examination is perceived as a necessary evil because of how it is carried out in practice: the questions posed may be unintelligible, or they may be ambiguous, or they may be highly speeded, or the student may be denied the opportunity of knowing what his performances are in the light of expected responses. Moreover, as hinted earlier, it may be the teacher has created an atmosphere which overemphasizes the assessment function. This may be the case when he deprives the student of the opportunity to have back the examination papers because they must be kept secure for use with other sets of students. This practice added to other undesirable elements bias the student's attitude against examinations.

The position taken here is that for those who think the examination system may be retained, the problem of improvement in use may be partly solved by emphasizing the learning function of classroom testing. This change of emphasis is in accord with the teachers professional role in the learning situation. Besides the new emphasis may hopefully change the student's perceptions of examinations, and the twin practice of grading. The evil aspects of the system will thus be minimized and conditions set for higher attainment of objectives--higher than the attainment possible under conditions where the assessment function is emphasized at the expense of the learning function.

If, for example, students take an examination in a classroom situation under the so-called "examination conditions"

and in addition repeat the examination "at home," making use of all available resources, excluding the teacher and fellow students, and their performances on both occasions count for their grades, then they may perceive the learning function of examinations. In this case, a student would be "cheating" if he were to solicit help from the teacher or his fellow student, within the "examination period." Other genuine efforts to seek out the correct response would then be encouraged and rewarded.

If, in addition, the students are made to "grade" their own performance to the best of their knowledge, they would be learning still in carrying out such a requirement, and they may grow to perceive and appreciate the meaning of examination grades. Classroom achievement examinations administered in this way may be described as improvement-in-use. From the point of view of both the teacher and the learner the modified practice neither eliminates nor depreciates the assessment function; but it pushes to the fore the learning function, and this is likely to be richly rewarding.

The specific problem of this study was to investigate these suppositions using students enrolled in two courses in the College of Education, Michigan State University. Further details about the students and the courses are given in Chapter 2. The main purpose at this stage is to introduce the problem both in general, and with specific reference to the particular conditions in which it was investigated.

How may achievement examinations be improved in use when applied to the courses selected, so as to better define and attain the objectives of these courses? To solve this problem examinations were administered twice each--in class, and outside class--to groups of students. Members of one of the groups were also required to evaluate their performances. Was such a procedure any improvement-in-use of examinations? The answer to this question will be found in Chapters 3 and 4. Meanwhile it will be necessary to relate such a practice to similar ones in the literature.

Related Literature

The problem posed and the solution proposed stem not only from a practical situation but also from two types of previous research studies. One type deals directly with the learning function of examinations and the other on the effect of knowledge of results and encouraging comments on student's examination performances. A few of these will be quoted to illustrate the connection.

Examinations as a Learning Device

In a study on "Examination as an Aid to Learning" Jersild (1929) sought to answer this question: to what extent does the examination enforce an active participating attitude of mind on the learner, and does such activity yield higher returns in achievement when compared to the attainment resulting

from ordinary conditions of study? He used two equivalent groups in each of a set of replicated experiments where the main treatment variable was what the author called "pre-examination." By this he exposed the "experimental" groups to an examination experience before using the same test items or constructing new ones to assess the groups' achievement of course objectives. Thus the experimental groups had examination "warm-up" during the pre-testing or "pre-examination" period. The other treatment variable was the examination-type; there were three types: true-false, multiple choice and essay.

There were five replications of this study. In the first two the "pre-examination" was made up of true-false items; multiple-choice items were used in the third and fourth experiments and the essay in the last one. Jersild's study is very relevant to the present one; three of the replications will therefore be described in some detail.

The first experiment, like the others, was carried out in a psychology class. There were two sections in the class, each made up of 37 students. The course objectives are stated in general terms to include an understanding of "classroom lectures" and "reading assignments." At the beginning of a semester one of the groups was randomly selected as the experimental group and given a "pre-examination" on materials to be covered in the next six weeks; the other class had not this treatment. At the end of the first six weeks both

classes were administered the same true-false examination used in pre-testing the experimental group.

The classes exchanged roles in the second part of the semester such that the one that served as the control, formerly, became the "experimental" group and was pre-tested on materials to be covered in the rest of the semester. In the end both groups were assessed on their achievement by the same true-false test used in pretesting the experimental group.

The third experiment also involved two classes, each with 42 students. The topic to be learnt was "Reaction Time." The experimental group was selected randomly and counter-balanced as described above. The "pre-examination" in this case was made up of multiple-choice items. But the final achievement was tested by newly constructed true-false and recall items.

The procedure in the last experiment (N = 63 in each group) followed the lines already described. But here the "pre-examination" for the experimental groups was of the essay-type, and the subject to be learnt was a biographical selection. A test of immediate recall was administered as a criterion measure.

The author summarizes the results of these experiments in the form of ratio scores $\frac{100(Me)}{M_C}$ where Me equals the mean score for the experimental group and M_C the mean score for the control. With the exception of replications in which

true-false items were used in the "pre-examination" the experimental group always scored higher than the control.

The study may be criticized on the ground that it did not control for the memory factor. But as most of the results were in the predicted direction one cannot reject completely the author's conclusion that the treatment group excelled the control in subsequent performance, and that the treatment stimulated "the industry of the learner."

The present study will use and modify Jersild's method of repeating the same examination with the treatment group; but the repeat will be outside class so that not only will the industry of the learner be challenged but also will he be able to use the examination directly as a study guide.

Standlee et al. (1960) investigated "quizzes" and their contribution to learning. The quizzes were made up of twenty true-false items and given at the end of each month of work; thirteen of them were administered during the experiment. There were three experimental conditions and a control. In condition 1 the quizzes were administered in the written form, graded by the instructor and the scripts were returned to the subjects. The author explained that the mere giving of quizzes would enforce the students' learning activity as well as provide a structuring of the course for the guidance of the students. The instructor's written grades provided extrinsic motivation; moreover the students had knowledge of their performance item by item as the corrected scripts were returned to them.

The second experimental condition received the quizzes in written form too; but the members checked their own work, presumably from key provided by the instructor. This group therefore experienced the same benefits as stated for those in condition 1, but without the extrinsic motivation from teacher-awarded grades. In the third condition the same quizzes were read out orally by the instructor who also provided the correct answers. The only benefit enjoyed by this group was the enforced activity and course structuring. The control group enjoyed none of these benefits as it had no quizzes.

All groups had a preliminary pre-test comprising 100 multiple-choice items which had been tried out in the same course in a previous semester. The scores on this were used as covariates in the analysis of the results. The criterion measures comprised of 100-item multiple-choice examination given at mid-semester, and a 150-item test given at the end. The mid-semester examination included 50 items from the pre-test while the final included the other 50 items which were in the pre-test, but not in the mid-semester examination. When the mid-semester scores were used as criterion, significant difference was found at the .06 level as against the .05 hypothesized. A t-test comparison of the means of condition 1 and the control was also significant at the .05 level. The differences were not significant with the finals as criterion.

It appears that the author's criterion measures were not sensitive enough, since they contained from a third to a half of the items on the pre-test. Furthermore, a "multiple comparison" technique like Scheffe's (1959) could have been used. It must be remembered also that the quizzes were made of true-false items, which according to Jersild (op. cit.) are of "dubious value as a pedagogical instrument." These limitations may have eclipsed the effect of treatment.

The present study also uses examination as the main treatment variable. But all the defects listed above are avoided. Moreover, the idea of the subject grading his own work is adopted and given much weight and significance in that the subject was given the opportunity to compare his self-evaluations with the evaluations from the teacher-experimenter.

In a similar study Fitch et al. (1951) investigated the effect of "frequent testing as a motivating factor in large lecture classes." The authors found that frequent testing resulted in "superior achievement." But they remarked that the "instructional function (is) best served when divorced from the regular process of achievement evaluation." The present study specifically challenges Fitch's (et al.) remark. The subjects were told at the beginning that all examinations administered would count towards their final evaluations.

Teacher's Comments

The studies hitherto mentioned were conducted in a college setting. Page (1958) couched his in a secondary school

setting. He had 74 teachers of different subjects from different schools involved in an experiment in which the treatment consisted only of "teacher's comments" on objective examination answer scripts. The subjects for the experiment were drawn from the 7th through the 12th grades and twelve schools were represented. The treatment variable was at three levels--"no comment," "free comment" and "specified comment" and subjects were assigned to treatment at random. The experiment basically involved administration of the treatment on the answer scripts of a first test and then using the performance on a second as criterion. Since a factoral design was used, the experimenter was able to investigate interactions between treatment and schools, and classes and school year. The results were analyzed by the Friedman Rank-Test and the effect of treatment was highly significant. The author concludes: "When the average secondary teacher takes the time and trouble to write comments . . . these apparently have a measurable and potent effect upon student effort . . . or whatever it is which causes learning to improve."

It would be interesting and valuable to know whether similar conclusions can be reached if the study were conducted in a college setting. The present study incorporated "comments" in one of the conditions. But perhaps its greatest connection with Page's work will be in the use of different courses, and similar test statistics in result analysis.

The Purposes and Hypotheses

The related studies reviewed provide evidence that classroom achievement examinations perform some learning functions besides their measuring function. The present study was not concerned with establishing additional evidence for this learning function; this was, and is assumed. Rather it was concerned with manipulating the examination variable in order to realize and increase its value as a learning device. As indicated earlier there is some dissatisfaction with examinations. Such a state of affairs would appear to result from the way examinations are operated, and not through any intrinsic attribute of examinations. One may even suspect that those who speak of the "tyranny of testing" would not gainsay its potentiality to stimulate and promote learning in a classroom situation.

Purposes

It was suggested earlier also that attainment of course objectives may be increased and that student attitude may become favorable towards examination and grading if the learning function is deliberately emphasized. The primary purpose of this study was therefore to spell out and test a method by which the learning function of examinations may be increased. This method consisted of administering two examinations within a term, followed by a final examination at the end of the term. The results of all three examinations count for the course

grade earned by each student. The two within-term examinations were manipulated as follows: the examinations were first taken by all students, that is taken under classrcom "examination conditions." Two randomly formed experimental groups then repeated the examinations in a home situation, and members of one of the groups were further required to evaluate and grade their performances. It was hoped that such "treatment" challenges the industry of the student and emphasizes to him the learning function of examinations and the meaning and significance of grading. This modification in examination procedure seeks to incorporate deliberately those practices judged to have high value as a learning device; moreover, it makes the student experience the problem of awarding grades.

A secondary purpose of the study was to survey and describe the attitude of the students involved towards examinations and grading. The need for a valid instrument to carry out such a survey defined another purpose: to develop an attitude scale battery.

Hypotheses

The following were the hypotheses emanating from the purposes just stated:

1) The experimental groups which repeated earlier examinations would score higher than the control group on subsequent examinations in the same course.

2) The experimental group whose members both repeated and evaluated their performances in earlier examinations would score higher than all other groups in subsequent examinations in the same course.

3) The attitudes of the experimental groups as measured by a specially developed scale would be more favorable towards examinations and grading than the attitudes of the control group.

The first two were tested at the .05 level of significance, but the results on the last hypothesis were used for rank-ordering the treatment groups on the criterion measure.

CHAPTER II

METHODOLOGY OF THE MAIN STUDY

By now it should have been obvious to the reader that the "treatment variable" in this study was the method of examination. It was the "variable" not in the quantitative, but in the qualitative sense. But it is necessary to explain how it was supposed to operate.

The Treatment

There were three conditions as follows:

1) an examination was taken under normal conditions and members of the group were required to repeat their performance in non-examination conditions;

2) an examination was taken under normal conditions and members of the group were required both to repeat and to evaluate their performances on the two occasions;

3) an examination was taken under normal conditions and members of the group were required neither to repeat nor to evaluate their performances; but they were permitted to keep the examination scripts as their property.

These three treatments may be referred to as T_1, T_2 and T_3, respectively. The requirements stated above are clear; but the second one for T_2 may easily be confused with the so-called "level of aspiration" type of experiment. In the latter, the subject "estimates" his score, for example; by and large

such estimates are guesses, depending as they do on past experiences of success or failure. It must be emphasized that the self-evaluations envisaged here should not be guessed estimates; if they are guesses depending on what "self-concept" the subject holds, then they are not the treatment implied in this study.

If the self-evaluations were not to be guessed estimates, what should they be? They were and should be scores and grades which the subject awards himself--solidly based on knowledge--present knowledge, which he gains by expending effort to use all possible resources, excluding the teacher and fellow students, to search out for the correct responses for the test items. In a University setting the requirement to carry out such a search is not beyond the student. Even in a High School setting with fairly adequate library and other learning facilities the student can cope with this requirement.

Is T_1 different from T_2? Both require effort to search out the correct answers. It is, however, claimed that the additional requirement imposed for members of the second group to judge their work induces them to pay more attention than do the members of the other group; should this be so they would also learn more. By the same argument members of T_3 might not learn as much as those of the other two groups.

It should be added that all the three treatment conditions emphasize the learning function of examinations.

Clearly, the "practice effect" is double for T_1, and T_2, and all three groups have the opportunity to use the test items as a study guide.

The Criterion Measures

The final examinations at the end of term reflect the course objectives; scores on these were used to test the hypotheses on students achievement.

The second set of criterion measures were scores on the "Students' Attitudes Towards Examinations and Grading" scale battery--a scale which may be called for short "SATEG" scale battery. This instrument was specially developed for this study. A brief account of the operations involved is relavant here.

Free response statements of opinions on examinations and grading were first obtained from a small sample of students through an open-ended questionnaire. These responses were content analyzed in a search for "significant" statements which focus on clearly specified attributes of examinations and grading. The selected significant statements formed nuclei for the initial sixty-five attitude statements constructed. These were rated and Q-sorted by ten judges. The forty and eight statements which survived that exercise were administered to a representative sample of students, and the responses were factor analyzed. Finally thirty-two items were selected largely on the basis of their high loadings on

the various "factors" revealed. There was therefore sufficient evidence both in the operations outlined and in the reliabilities of the factor sub-scales--enough evidence, that is, to show that the scale is fairly valid for the purpose for which it was designed. Full details of the operations will be found in Appendix A.

The Population, Sampling, and Allocation to Treatment

The population used in this study was made up of students enrolled in two courses during the Fall Term, 1968. The courses are (1) ED 200: Individual and School and (2) ED 450: School and Society. Both are offered in the College of Education, Michigan State University.

In ED 200 there were 34 classes, each made up of at least 30 students. Sixteen instructors were in charge of 32 classes--one each in the morning and one each in the afternoon. Two other instructors were in charge of the other two classes, one for each. In one of these another experiment was in progress, and to control for possible contamination from this source the class was not sampled; the other class was also withdrawn since its instructor had one class and not two as the others. Thus sixteen instructors class-groups were left for sampling. Fifteen of these were randomly selected, randomly formed into three equal groups and the groups then randomly assigned to the experimental conditions. The selection, formation of grcups and allocation to treatment was done separately, and based on a table of random numbers.

In ED 450 there were twelve classes of at least fourteen students each. These were under seven instructors, five of whom taught two classes each. The other two had one class each. All classes here were involved in the study. These twelve classes were randomly formed into three groups and the latter were then randomly allocated to the three treatment conditions.

The population thus defined is rather limited and conclusions will largely be confined to it. But it may be argued that it represents typical education students as these two courses are required of all education majors. To the extent that these students are typical of education majors in particular and college students in general, the conclusions may be extended. However, no attempt at such wide generalization will be made from this study--as yet.

The Design

The main elements of the design have been described, but it is necessary to add that the study was conducted as two separate experiments. The one in ED 200 will be referred to as experiment 1, and the one in ED 450 as experiment 2. The resulting sub-designs are illustrated in tabular forms below:

SUMMARY OF SUB-DESIGN FOR EXPERIMENT 1

T_1					T_2					T_3				
C_1	C_2	C_3	C_4	C_5	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{21}	C_{22}	C_{23}	C_{24}	C_{25}
C_6	C_7	C_8	C_9	C_{10}	C_{16}	C_{17}	C_{18}	C_{19}	C_{20}	C_{26}	C_{27}	C_{28}	C_{29}	C_{30}

SUMMARY OF SUB-DESIGN FOR EXPERIMENT 2

T_1		T_2		T_3	
C_1	C_2	C_5	C_6	C_9	C_{10}
C_3	C_4	C_7	C_8	C_{11}	C_{12}

KEY: T = Treatment
C = Class (nested within Treatment)

Procedure

This study was a practical classroom experiment. It is therefore appropriate to describe first how the courses used are normally organized, and then the execution of the experiment and how it was woven into the existing structure.

There is always a large enrollment in the two courses. In the period of study, the totals were 1129 and 185 for ED 200 and ED 450 respectively. The lectures are given by a team of professors including the Course Coordinators who are also responsible for all arrangements relating to the courses. The students are divided into "discussion" groups under the leadership of graduate assistants, as instructors. These

groups constituted the "classes" which were the experimental units in the present study.

The administrative operations were conducted at three levels--(1) arrangements with the Course Coordinators, (2) contact with the instructors and (3) students' activities.

Consultation with the Course Coordinators preceded and continued throughout and beyond the study period. They were informed of the nature and purpose of the study through discussion, and the proposal was made available to them. They in turn supplied the writer information on the number of discussion groups and their instructors. The latter formed the basis for the definition of classes, formation of groups and allocation to treatment. All these were done randomly as described earlier. Furthermore, the Coordinators were told in discussion and in writing the type of scores that would be used as criterion measures. Such information would be kept in their records which would be made accessible to the writer when he needed them.

Contact with the instructors was to be kept at a minimum. There were reasons for this. First, the writer did not wish to bias any of them for or against the treatment; secondly he would have preferred an atmosphere in which no fuss about the study existed, and in which as far as possible the subjects remained naive; thirdly, it was desired to see how far the procedure for carrying out this study could be understood from written instructions only. More will be said on these points

in chapters IV and V. Meanwhile, it will suffice to say that the instructors were expected to follow written instructions but that in actual practice the writer dealt with problems individually as they arose. These were very few in the T_1 condition. Most of the problems arose in connection with T_2 making it necessary to eliminate certain aspects of it. Originally the members of this group were expected to graph their scores and grades. Such a graph was called a "progress chart" and was to be submitted to the instructor for "comments." Furthermore the instructor was to allow at least ten percent of his assigned grade to the activities involved in this study. These aspects of the treatment were eliminated because they involved both the student and the instructor in too much work.

Appendix B presents all that was originally designed for both treatment conditions and includes the supplementary instructions in full. Here it is only appropriate to present the instructions as they actually applied. These were given orally by the instructor, and woven into his planned activities for his class.

Instructions for T_1 Condition

1) "You will be expected to repeat each of the two within-term examinations at home. You may take up to four days before submitting this second attempt for scoring."

2) "You will be free to make use of all resources, excluding instructors and fellow students. Your aim should be to come out with all answers correct, working independently."

Instructions for T_2 Condition

1) "You will be expected to repeat each of the two within-term examinations at home. You will be free to make use of all resources, excluding instructors and fellow students. Your aim should be to come out with all answers correct, working independently."

2) "You will be expected to score and grade your two performances. Score, using your best judgments on what you feel are the correct answers. Evaluate your scores by assigning grades to yourself (0 - 4.5) using some criteria you feel to be objective."

3) "You may take up to four days before submitting your second performance for machine scoring."

4) "Later when you receive the feedback, check your scoring and self-evaluation and discuss the discrepancies with your instructor, until you are satisfied."

During the study it became necessary to issue supplementary instructions for this group. They were likewise addressed to the instructor. Here again the full instructions will be found in Appendix B (c). The relevant portions actually adopted were as follow:

1) "Ask your students to:
   a) write their names on their test booklets--to help them recover their copies,
   b) mark their in-class performance on both the test booklet and the answer sheets provided; the answer sheets will be handed in but they will keep (or pick up later) their test booklets to score and grade the markings at home--as described below."

2) "Give to every student a spare answer sheet and a pencil for the repeat performance described below."

3) "Emphasize that every student is to rework the test making use of all possible resources, excluding fellow students and instructors. To prevent any embarrassment over wide discrepancies this exercise must be done first and with care."

4) "When and only when the student has established enough confidence in his/her answers on the second performance (without any consideration of the first), then

and only then should he/she proceed to score and grade this second performance. Emphasize that guessing in any form will result in wide discrepancies."

5) "With the scoring and grading of his/her repeat performance as the "Key" the student then turns over to his marked test booklet to score and grade that performance also.

Treatment Condition T_3

This was the "control" group; the members were allowed to keep their test booklets, but no other requirements were expressed.

The third level of operation may be described under students' activities. These consisted of their following instructions as these were communicated to them through their instructors. Members of both groups T_1 and T_2 repeated their performances in the examinations and re-submitted their work for machine scoring. But in some classes, and particularly in experiment 2 there were misinterpretations of the self-evaluation requirement at the beginning of the experiment. As mentioned earlier it became necessary to issue supplementary instructions; after that there were no more problems.

Members of the control group (T_3) were not required to do anything other than take back the examination scripts which they kept as their properties. Finally, it is also relevant to note that all students were given written "keys" to the two within-term examinations. These were however delayed for about four days until members of groups T_1 and T_2 had turned in their second performances and self evaluations. Instructors also discussed the tests in class.

Analysis

The results of this study were analyzed by nonparametric methods. In particular the Friedman X_r^2 was used to test the overall effect of treatment with respect to the hypotheses on achievement of course objectives. This was followed by a t-test comparison of group mean scores. The groups were ranked on the attitude criterion according to their mean scores and percent of high scorers on each sub-scale. Full details of this analysis and the outcome are presented in chapter III.

CHAPTER III

RESULTS AND ANALYSIS

The use to which the results of the first mid-term examinations were put is given in this chapter. This is followed by the outcome of the study with respect to the hypotheses investigated. The analysis is made for each of the two experiments separately.

Experiment 1

There were ten classes under each treatment condition in experiment 1. Their mean scores on the first mid-term examination are shown in Table 1 on the following page.

It will be noticed as one reads down the columns under each treatment condition that the scores are arranged in descending order of magnitude. Thus Class 1 occupies top rank position within the T_1 group; class 11 occupies top rank position within group T_2; similarly class 21 is top in the T_3 group. As a further illustration class 9 is ninth in T_1; class 19 and class 29 are also ninth in the groups T_2 and T_3 respectively. This arrangement makes it possible to match the classes according to their rank positions in their respective groups. It turned out as the table shows that the

TABLE 1

MEAN AND RANK SCORES ON FIRST EXAMINATION (EXPERIMENT 1)

T_1				T_2				T_3			
Class	N	Mean Score	Row* Rank	Class	N	Mean Score	Row* Rank	Class	N	Mean Score	Row* Rank
1	31	26.52	1	11	37	26.11	2	21	32	25.84	3
2	32	26.50	1	12	37	25.95	2	22	35	25.51	3
3	35	25.91	1	13	29	25.76	2	23	36	25.47	3
4	35	25.31	2	14	34	25.73	1	24	32	25.29	3
5	36	25.03	2	15	28	25.60	1	25	36	25.19	3
6	32	24.91	3	16	35	25.43	1	26	33	25.15	2
7	39	24.36	3	17	43	25.16	1	27	34	24.97	2
8	35	24.26	3	18	35	24.86	2	28	36	24.97	1
9	34	24.15	2	19	38	24.82	1	29	35	24.21	3
10	29	22.97	3	20	33	23.91	2	30	32	24.06	1
Group Mean Score		24.99				25.33				25.07	

*The highest has the rank 1.

matched classes would have very nearly identical scores if these were reduced to two significant figures.

The columns headed "Row Rank" reflect the absolute differences in the scores of members of the matched triples. Classes 1, 11, and 21, for example, have scores of 26.52, 26.11 and 25.84 respectively; their rank scores within the triple are therefore 1, 2 and 3 respectively. Scores for classes 7, 17 and 27 are 24.36, 25.16 and 24.97, and the corresponding rank scores are 3, 1 and 2 respectively. The Friedman test (Siegel, 1956) was applied to test the significance of the difference of the sum of ranks shown in the "Row Rank" column. This was not significant ($X_r^2 = 4.2$; $X^2_{r(.05)2} = 5.99$). Evidently the differences among the groups were not statistically significant at the start of the experiment.

The Hypotheses on Achievement of Course Objectives

The first hypothesis was that the mean score for group T_1 would exceed the one for group T_3 in the achievement of course objectives as measured by the course end examinations. The second hypothesis maintained that the mean for T_2 would exceed each of the means for T_1 and T_3. The final examination results presented in Table 2 on the following page were used in testing these hypotheses.

Table 2 shows that the classes in each matched triple were ranked on the basis of their mean scores, as illustrated earlier. The Friedman test was applied to test the overall

TABLE 2

MEAN AND RANK SCORES ON FINAL EXAMINATION (EXPERIMENT 1)

T_1			T_2			T_3		
Class	Mean Score	Row Rank	Class	Mean Score	Row Rank	Class	Mean Score	Row Rank
1	54.83	2	11	56.22	1	21	54.45	3
2	54.36	2	12	55.25	1	22	54.37	3
3	54.97	2	13	55.35	1	23	52.40	3
4	53.90	1	14	53.56	2	24	51.68	3
5	53.97	2	15	52.55	3	25	54.44	1
6	53.41	2	16	54.03	1	26	51.30	3
7	50.66	3	17	52.07	1	27	51.25	2
8	53.51	2	18	54.03	1	28	52.44	3
9	51.81	2	19	53.61	1	29	51.74	3
10	50.93	3	20	53.09	1	30	51.85	2
Treatment Mean Score	53.24			53.98			52.59	

treatment effect. The difference was significant at the .05 level ($X_r^2 = 8.6$; $X^2_{r(.05)^2} = 5.99$). This means that the risk involved in rejecting a contrary ("null") hypothesis that the treatment produced no effect has a probability of about five percent: in other words the probability is high that the null hypothesis is false. If so the alternative that there was a treatment effect may be accepted.

A t-test comparison was then made between the pair of means for T_1 and T_3. The difference was not significant ($t = 0.995$; $t_{.05(18)} = 1.734$). The meaning in this case is that the treatment effects on these two groups, if any, were not significantly different.

The second hypothesis was in two parts: part 1 involves comparison of the means for T_2 and T_1; these as the table shows are almost identical. The other part involves the means for T_2 and T_3. A t-test showed that the mean for T_2 was significantly larger than the one for T_3 at the .05 level as hypothesized ($t = 2.35$; $t_{.05(18)} = 1.734$). The chances are therefore small--about five percent--that the null hypothesis of equality of means for the two groups is true. The alternative experimental hypothesis was therefore accepted that the mean for T_2 was significantly larger than the one for T_3.

The nonparametric test revealed there was an overall significant treatment effect. Would a parametric test lead to such conclusion? To provide answer to this question the criterion scores were re-analyzed by the analysis of

co-variance method. The means for the first examination were used as co-variates. As mentioned earlier they were not significantly different, but the F-value of 2.08 suggested there might be one or two very large scores, so that it would be advantageous to remove the variance associated with initial test scores. Table 3 summarizes the results of this analysis. It is evident that the gain in the co-variance analysis is only slight. Without it the F-value is 2.45, and significant at 20% ($F_{.20(2,27)} = 1.71$); with it F is 2.76 and significant at 10% ($F_{.10(2,26)} = 2.52$). In neither case is the difference significant at the five per cent level, as was hypothesized.

TABLE 3

CO-VARIANCE-ANALYSIS OF THE FINAL EXAMINATION SCORES-Y (EXPERIMENT 1)

Source	SS_X	SS_{XY}	SS_Y	$SS_{Y'}$	df	$MS_{Y'}$	F
Between	0.643	1.915	9.593	5.901	2	2.951	2.76*
Within	18.036	21.252	52.860	27.818	26	1.069	
Total	18.679	23.167	62.453	33.719	28		

*Not significant; $F_{.05(2,26)} = 3.37$.

A similar covariance analysis of the means for T_2 and T_3 also revealed no "significant" difference; but the F value was .02 less than the one required for significance. The

tabular illustration below displays the relevant data.

COVARIANCE ANALYSIS OF THE FINAL EXAMINATION SCORES FOR T_2 vs. T_3 IN EXPERIMENT 1

Source	SS_X	SS_{XY}	SS_Y	$SS_{Y'}$	df	$MS_{Y'}$	F
Between	0.4565	1.8476	9.5773	5.5367	1	5.5367	4.4173
Within	6.6420	8.0139	30.9776	21.3085	17	1.2534	
Total	7.0985	9.8615	40.5549	26.8452	18		

$F_{.05(1,17)} = 4.43$; $F_{.10(1,17)} = 3.03$

Experiment 2

Table 4 on the following page shows the mean class scores on the first examination for the classes and groups in

TABLE 4

MEAN AND RANK SCORES ON FIRST EXAMINATION (EXPERIMENT 2)

T_1				T_2				T_3			
Class	N	Mean Score	Row Rank	Class	N	Mean Score	Row Rank	Class	N	Mean Score	Row Rank
1	17	37.94	1	5	14	36.71	3	9	16	37.44	2
2	14	36.86	1	6	15	36.33	2	10	18	36.17	3
3	14	36.86	1	7	17	35.77	3	11	14	35.86	2
4	19	35.47	2	8	14	34.71	3	12	13	35.69	1
Group Mean		36.78				35.88				36.24	

experiment 2. The ordering of the classes within each condition and their consequent matching and ranking within matched triples across treatment conditions were done exactly as described in experiment 1 earlier. The Friedman test was also applied to test the significance of the sum of ranks in the "Row Rank" column. The groups were not statistically different ($X_r^2 = 4.50$; $X_{r\,.05(2)}^2 = 5.99$).

As in experiment 1 the hypotheses were that

(i) the mean for T_1 is greater than the mean for T_3

(ii) the mean for T_2 is greater than the mean for T_1

(iii) the mean for T_2 is greater than the mean for T_3.

Table 5 below presents the data for testing these hypotheses. But it is clearly evident that there is no need to apply any statistical tests: the group means are almost identical, and the figures in the "Row Rank" column show a pattern contrary to the hypotheses.

The mean class scores on the final examination are shown in Table 5 below.

TABLE 5

MEAN AND RANK SCORES ON FINAL EXAMINATION (EXPERIMENT 2)

T_1			T_2			T_3		
Class	Mean Score	Row Rank	Class	Mean Score	Row Rank	Class	Mean Score	Row Rank
1	46.24	2	5	45.79	3	9	46.63	1
2	44.93	3	6	45.50	1	10	44.94	2
3	47.43	1	7	45.71	2	11	44.50	3
4	44.74	2	8	43.77	3	12	45.67	1
Group Mean	45.84			45.19			45.44	

The pattern shown in the above figures is contrary to the hypothesis. The differences between treatment conditions were, however, not significant ($X^2_{r\,.05(2)} = 5.99$). The risk of rejecting the null hypothesis in this case would be as high as 93 per cent (Siegel, Table N).

The Hypotheses on Attitudes

The third major hypothesis of the study was that the attitudes of the experimental groups would be more favorable towards examinations and grading than the attitudes of the control group. In view of the breakdown of the scales described in the Appendix A, and in view of the position taken there of the nature of attitude this hypothesis will be sub-divided and examined in parts and with reference to the attitude "factors". These sub-hypotheses are:

1) that <u>each</u> of the groups T_1 and T_2 would score <u>higher</u> on the "learning function" factors (EP and GP) than group T_3.

2) that <u>each</u> of the groups T_1 and T_2 would score <u>higher</u> on the "motivating function" factors (EP and GP) than group T_3.

3) that each of the groups T_1 and T_2 would score <u>lower</u> on the "Dys function" factors (EN and GN) than group T_3.

4) that <u>each</u> of the groups T_1 and T_2 would score <u>lower</u> on the Pressure-Anxiety factors (EN and GN) than group T_3.

The first two of these sub-hypotheses re-echo the parent hypothesis, and also specify the crucial attitude "anchors." The other two say the same things indirectly, since "lower" placement on the "negative" dimension is a "more favorable"

position, relatively. Table 6 presents the mean scores on the factor scales. The measure is the same in both experiments, hence the results are reported under each treatment condition, with the groups in the two experiments combined.

TABLE 6

MEAN GROUP SCORES ON THE ATTITUDE FACTOR SUB-SCALES

Scale	Factor	T_1 (N=265)	T_2 (C=219)	T_3 (N=245)
EP	Examination Satisfaction	1.62	1.61	1.62
	Learning Function	2.19	2.35	2.23
	Motivating Function	2.11	2.37	2.15
EN	Examination-type	2.92	2.79	2.97
	Dysfunction	3.04	2.84	2.96
	Pressure-Anxiety	3.66	3.55	3.57
	Hate	3.06	2.80	3.03
GP	Learning Function	2.27	2.40	2.30
	Motivating Function	2.69	2.88	2.66
	Measuring Function	2.62	2.76	2.57
GN	Dysfunction	3.24	3.31	3.23
	Pressure-Anxiety	3.71	3.74	3.75
	Hate	2.71	2.59	2.67
	Non-learning	3.27	3.00	3.25
	Non-measuring	3.56	3.41	3.41

The absolute scores shown on the above table are so close that they may not be significantly different; but ranking across treatment conditions (1 for the highest) produces the following pattern of rank scores for the crucial factors specified in the sub-hypotheses:

		T_1	T_2	T_3
EP	Learning Function	3	1	2
	Motivating Function	3	1	2
GP	Learning Function	3	1	2
	Motivating Function	2	1	3
EN	Dysfunction	1	3	2
	Pressure-Anxiety	1	3	2
GN	Dysfunction	2	1	3
	Pressure-Anxiety	3	2	1

When T_1 and T_3 are compared the trend shows T_1 scoring lower on the EP and GP factors and higher on the EN factors. This is contrary to expectation. On the other hand when T_2 and T_3 are compared T_2 scored higher on the EP and GP factors and lower on the EN factors. This fact tends to support the hypothesis. The pattern for the GN factors is not consistent.

The results above consider the means of the groups. The extreme scores throw further light on the relative positions of these groups on the attitude factors. The percents of the group choosing each point on the Likert Scale are given in Appendix C. An extract from that Table gives the following picture. On the learning function factor the percents of respondents choosing point 4 and 5 were 17, 13 and 14 for T_1, T_2 and T_3, respectively. It would be expected that more students in T_1 than in T_3 should be "high" on this factor. The trend is in line with this expectation. On the other hand comparison between T_2 and T_3 shows a contrary trend.

The trend is consistently in line with expectation when the groups are compared on the motivating factor. The corresponding percentages are 18.5, 16.5, and 14.5 for T_1, T_2 and T_3 respectively.

The dysfunction factor responses revealed the same pattern as the learning function factor. T_1 was lower than T_3 as would be expected; but T_2 was higher than T_3--against expectation. The respective figures are 27, 34 and 30. On the Pressure-Anxiety factor the trend falls in line with the expectation. The values are 51, 51 and 53 respectively. Table 7 converts the percentages given here into rank scores, and thus makes it easy to comprehend the relative positions at the "high" extreme end of the scale factors.

On the Grading Scales the trend was consistently in the opposite direction as illustrated by the following percentage figures:

	T_1	T_2	T_3
Learning function	17	18	18
Motivating function	31	30	33
Dysfunction	48	44	31
Pressure-Anxiety	65	67	65

These percentages are also converted to rank scores in Table 7 on the following page.

TABLE 7

RANK-SCORE POSITIONS AT THE HIGH EXTREME END OF THE FACTOR-SCALES

Factor Sub-Scales	T_1	T_2	T_3	Comments Relating to the Hypothesis
EP:				
Learning function	1	3	2	T_1 is in line with hypothesis T_2 is contrary
Motivating function	1	2	3	The pattern is consistent with the hypothesis
EN:				
Dysfunction	3	1	2	T_1 is in line: T_2 is contrary
Pressure-Anxiety	2.5	2.5	1	The pattern is consistent with the hypothesis
GP:				
Learning function	3	1.5	1.5	The patterns for the GP and GN factors are both inconsistent and generally contrary to the hypothesis
Motivating function	2	3	1	
GN:				
Dysfunction	1	2	3	
Pressure-Anxiety	2	1	3	

CHAPTER IV

DISCUSSION

Interpretation of the Results

This chapter is an overview of the results of the last one with specific reference to the hypotheses of the study. To what extent did these facts agree with the hypotheses, and what were the limitations? The chapter also examines some implications of the study, presents suggestions for further research and draws some tentative conclusions based on the present study.

The Achievement Hypotheses and the Statistical Tests

As reported in the last chapter the overall treatment effect was tested by the use of the Friedman X_r^2--a nonparametric method. The first showed that in experiment 1 there was a significant difference among the treatment conditions at the hypothesized level of five percent. But the second did not confirm such a finding.

Moreover the absolute means for the treatment groups were 53.24, 53.98 and 52.59 for T_1, T_2 and T_3 respectively. These are so close that the evidence for a treatment effect was very weak indeed; but the trend was in the predicted

direction. It should be remembered that the "control" group was not properly controlled as its members also had the opportunity to keep their examination test scripts. In these circumstances some, at least, of the members might have used the previous examinations as study guides.

Another objection to a possible conclusion that there was a treatment effect was the fact that the results of experiment 2 did not warrant such a conclusion. But the conditions in the course used in this experiment was rather peculiar. There was an open expression of "concern" by members of the control group that the other groups were being placed on an advantage by being required to repeat the examinations outside class. This concern "worried" the Course Coordinator. Once again this was admittedly the problem of Control; it was weak. But the existence of this concern openly would lead one to suppose that some members of the Control group were making use of the previous examinations as study guide.

Furthermore in T_2--the group whose members both repeated and evaluated their own performances--the treatment was misapplied in the first examination. The instructions were interpreted as if this was the so-called "level of aspiration" experiment. As was emphasized earlier (see chapter 2) this was not the case. The result of this misapplication was that discrepancies between the students' self-evaluations and the instructors' turned out to be so wide as to provide another source of concern, this time for one of the experimental

groups. The effect on members interest and attitude towards the whole exercise cannot be determined; it may be it is reflected in the low mean scores.

Granted that these observations were the facts operating within experiment 2, its results may not be taken as a reliable evidence against the existence of an overall treatment effect. But the results of experiment 1 were in the predicted direction. The nonparametric test was significant at the hypothesized level and the covariance analysis showed the risk of rejecting the null hypothesis not higher than ten percent. Even so no definite conclusion could be made on the overall effect of treatment if the two experiments were considered together: their results were conflicting.

It is necessary to elaborate the statement just made in the context of the two relevant hypotheses. As pointed out in the previous chapter the mean achievement score for T_1 (the group that merely repeated the examinations) was not significantly larger than the mean for T_3 (the "control") as was hypothesized. The point has already been made that both groups might in effect have been experiencing the treatment.

The second hypothesis stipulated that the mean for T_2 would exceed each of the means for T_1 and T_3. The results for experiment 1 and the t-test analysis show that this hypothesis was not supported when T_2 and T_1 were compared; but between T_2 and T_3 the difference was significant at the

five percent level. The covariance analysis in this case showed a risk considerably less than ten percent if the null hypothesis were rejected. Undoubtedly T_2 was the source of the weak tendency towards the predicted direction evident in experiment 1. The obvious question is: how was this treatment supposed to work?

In the first place the treatment which members of T_2 experienced would stimulate much effort if it would make the students spend extra hours of work to find out the correct answers. According to Jersild (op. cit.) examinations stimulate "the industry of the learner." It may be argued that in such a case it was the extra hours of work that led to increased learning and increased achievement, and not the repetition and self-evaluations as such. No one would deny that sheer effort as expressed in the extra hours of work is one of the significant factors explaining these results. But in the circumstances of this study work effort was a secondary factor, and brought about by the treatment condition--a specific requirement to exert the effort. It would then follow that the primary factor was the imposed treatment condition. Viewed this way, work effort is not a contaminating factor but a necessary secondary factor serving as the medium through which the primary factor operated.

The second element in the treatment was that the group had opportunity to use the test items as a study guide. Assuming that the examinations were made up of valid test

items, then they would be of immense value to guide the learner to the types of skills considered essential. They would also be of immense value in teaching the learner "test-wiseness" with respect to the language of test items and other test characteristics. An "examination set" to quote Meyer (1936) is thus developed and this would influence the learners methods of study.

A fourth means by which the treatment was supposed to produce the effect was the double practice session which the exercise brought about. Stone (1955) argues that the mere taking of an examination provides a "practice session" which aids "the fixation of correct responses . . . and the elimination of error." If this be granted, then the treatment as defined here would provide a reinforced practice session and would tend to promote learning on that count, per se. Overlearning may not be detrimental to learning.

Another element in the treatment was students' self-evaluation. Here again if one accepts Stones peculiar definition of "reinforcement" as "the fixation of correct and elimination of erroneous information," then self-evaluation would be a sort of "reinforcement." Moreover, it would tend to aid the development of critical ability which was further sharpened when the student discussed his "discrepancies" with the instructor. The discussion of the discrepancies and their resolution would further lead the student to perceive the meaning and proper function of examinations and thus help to motivate him.

These modes of operation of the treatment may be re-iterated for emphasis. It would stimulate extra work effort, serve as study guide to direct the learner, be valuable as an indirect way of teaching test-wiseness and developing a useful examination set. The additional practice session would provide a necessary and not a superfluous reinforcement. Add to these the advantages of self-evaluation which may include the development of critical ability besides the reinforcement opportunities it would provide. If these and other claims existed they would produce the type of effect which was evident, though very weak, in the T_2 condition in experiment 1.

The second tentative conclusion of this study therefore re-echos parts of the second major hypothesis. It is this: if a classroom achievement examination is administered under the conditions defined for T_2 of this study, if, that is, the student is given the opportunity to keep the test script, to rework it outside class, and to evaluate his performances there would be a tendency for his learning (and achievement) to rise; it would be raised higher than what it would have been if such conditions did not exist along with other devices in the learning situation.

Admittedly this treatment requires expended effort, but it is guided effort in the right direction. In fact some students spend much more effort than necessary and end with very little or no gain because they head in the wrong direction.

The Treatment Effect on Attitudes

The results of the attitude measures are anything but definitive with respect to the hypothesis that the experimental group subjects would show a more favorable attitude towards examinations and grading than members of the control group. As Tables 6 and 7 show the positions of the groups did not always agree with the hypothesis. Moreover, the results were in most cases not consistent when the basis for comparison was the group mean score, or when it was the percentage of high scorers on the factor scales. No conclusions could be made from such haphazard results.

What were the limitations in this case? The anchors of attitude for examination and grading are not located within the confines of one particular course, but within the total environment of the University setting. If so, it would require a longer period of treatment within a wider range of situations in the University for the treatment "to work," assuming that it was potentially effective. The point to be emphasized is that no conclusions could be made on the "effects" of the treatment on attitudes; nor could any be made on the alternative position of "no-treatment effect."

It may be added that there is some evidence of construct validity of the scales from the results as shown in Tables 6 and 7. In line with the position taken of the nature of attitudes towards examinations and grading it would be expected that if the student is "high" on the learning function

and motivating function factors then he should be low relatively on the dysfunction and pressure-anxiety factors. It is interesting to note that whether the basis of comparison is the group mean scores or the percentage of "high" extreme scores, the pattern of rank scores consistently tend to support this position. The only conclusion that may be drawn from those results is that they did tap the attitudes as defined in this study, despite the limitations of "response sets." These are supposed tendencies of subjects to respond to test items in some stereotyped manner; for example, a subject may reveal a set to respond "true" in a true-false test, another may show a tendency to prefer the middle position on a rating scale. The effects of such set tendencies could not be determined in the present study; however, attempts were made to minimize them in the peculiar way in which the Likert points were defined.

Implications of the Study

This was supposed to be a classroom experiment. How feasible may the "treatment" conditions described here be applied in practice in a normal classroom situation? This question may be broken into two parts--from the point of view of both the student and the teacher. The answers essayed will hopefully clear some of the doubts which may exist concerning the feasibility of the treatment. Moreover, the problems discussed here will help to bring together what further studies may be necessary to supply additional evidence on the present issue.

Questions and Answers For the Student-Critic

It may be asked on behalf of the student: what other values--apart from the learning function--may accrue from the treatment? The self-evaluation treatment may combat the tendency to guess wildly; it may also emphasize objectivity in the development of the so-called "self-concept." An incident that occurred during the study illustrates the latter point. A student broke down in tears because of the wide discrepancy depicted by her graph of "self-evaluations" as compared to the instructors evaluations. As mentioned earlier this aspect of the treatment had to be deleted to save the students (and instructors) from further embarrassments. It is the writers view that such incidents could be utilized to emphasize the need to be objective, to be realistic in making "self-evaluations." In other words, the treatment potentially has a "mental hygiene" value, and this may be exploited in an actual classroom situation.

Could the time involved in carrying out the "treatment" not be more profitably used in extra reading and other assignments to widen the students knowledge in the course? This is a very important question, with an important omission: the treatment is in effect an extra reading assignment. Any student who has ever done a "take-home" examination knows what extra reading he has to do on his own to produce an answer. The "treatment" specifically asks the student "to make use of all resources, excluding instructors and fellow students."

The incidental learning, to a student who experiences such a "treatment" may be as profitable as, if not more than the learning from extra reading assigned by the teacher, without the treatment. This statement assumes that the examination is made up of good valid test items with reference to specific course objectives.

The critic may still ask: will the student be able to cope with the amount of work involved if the "treatment" were applied in all the courses he takes? There is no simple answer to this question. It will depend upon the institution's educational objectives and policies for students' course loads; it will also depend on the capacities of individual students. Whatever the case it would not be too difficult for the student to adjust his course load, should the treatment be adopted as a general practice. Of course no such adoption is warranted by the weak evidence provided by this study alone.

Questions and Answers For the Teacher-Critic

There are at least two problems of serious concern to the teacher. He may wish to know what time involvement this treatment would demand of him. Secondly, he may be uneasy about having to lose good items in the process.

The second treatment condition even in its deleted form required the instructors to discuss students "discrepancies" individually, as they were brought up by each student. No estimates of the actual times involved could be made as these would vary widely according to the individual problems.

But it would certainly be "a lot of time"--to quote one instructor, if the class is large. In experiment 1 each instructor was in charge of at least sixty students. It must be remembered that these instructors were themselves students; one can then appreciate why some of them felt that the treatment was too demanding of their time. The burden may not be felt so much in a class about half that size, and in charge of full-time teachers. In any case, the burden of extra time involvement has to be weighed against the potential benefits of individualized instruction which this treatment also fosters.

The last remark also applies to the teacher's concern over loss of items. The crux of this study is on methods to increase the learning value of test items. Test items have learning value only if they are good valid items, and if they are, then there may not be a serious loss if the "treatment" would help to bring about maximum learning, which is the goal of instruction.

Another point is that course requirements are not static in a dynamic educational system. Nor are successive groups of students identical though they may be assumed to be from the same population of students. In view of these considerations a teacher may not use old test items without remodeling them as it were.

A study conducted by Ebel (1968) and to be published soon is relevant on this issue. The author was investigating the relative effectiveness of "new" and "old" test items as

assessment tools in achievement testing. The "new" items were newly constructed but the "old" ones were taken directly from the authors supplementary text to his book (Ebel, 1965). Students had previously been informed that some test items would be drawn from the supplement. The results of this study showed that the "new" and "old" items were quite comparable in discrimination index as well as difficulty values. These values however tended to be lower for the "old" items, but the correlation between performance on both types of items was positive and high--.64 in one case and .51 in another. The author concluded: "it is feasible to use some of the "old" items to measure achievement." Should such results be collaborated by replicated studies the teacher need not feel too concerned over loss of items. He can still keep files of old test items and re-use "some" of them--remodeling them in line with the dynamics of his subject.

Suggestions for Further Study

The conclusions of this study would still have been tentative--even if the results had been significant at the .0001 level. One of the main objectives of the discussion hitherto is to present evidence to show that the "treatment" is worth reconsideration. The practical problems involved are so slight that one cannot help suggesting and urging that this study be repeated in many and varied situations. Of course, in such replications the weaknesses of the present one must be avoided.

Two of these weaknesses deserve special mention. The first is on teacher-experimenter involvement and the second on the "control" group. For reasons stated earlier contact with instructors was forced to a minimum at the early part of the study. If this is to be a co-operative study by teachers, a study in which they are seeking ways to derive maximum possible value from achievement testing in the classroom, then they need to be told in advance the details and purposes of the study. This will hopefully get them involved. The "treatment" is believed to be such that it is experienced individually by the student and there may not be much contamination of results if teachers know about all the other levels of the treatment, provided they do not require any more of their student subjects than their assigned treatment. In a declared co-operative study the teachers will satisfy this proviso.

The other main weakness was to be found in the control group. It would be absolutely necessary to ensure a proper control group whose members are not exposed to any part of the "treatment." The writer believes this was the main source of contamination in the present study. To get over this problem it is suggested the experimental "treatment" conditions may be carried out in one term (or semester) and the control condition in a subsequent one. The problems of group equivalence would not be insurmountable. In such a case the course requirements, teachers and examinations used would remain the same for both groups.

Furthermore, it is suggested that many Departments of the University may be included in the study. In fact, at a later stage different Universities may be incorporated, provided they use the examination technique and base their grading decisions largely on it. It sounds an enormous venture. But if the present one may be considered as a pilot study the experience is that it is quite feasible. Each teacher within each department within each University will be conducting his own study, if all concerned agree to take part in the co-operative study. The reports of each experiment would be submitted to a co-ordinator who will extract the relevant data for a final analysis. The problem of different criterion examinations would be easily handled by following the lead of Page (1958) and converting raw scores to rank-scores and using appropriate non-parametric tests as illustrated in this study. Such an encompassing study would take one academic year to finish. It would be worth it.

The latter suggestion about running the study in a wider and more varied setting underlies one other limitation of the study. The setting was too narrow and the period of time too short to allow the "treatment" a fair chance to work on students' attitudes. The extended design will meet this limitation.

If this be taken as a pilot study the experience is that it is largely feasible and that the implications from the students' as well as the teachers' points of view can be

accommodated. If this is so, the study is worth repeating so that additional evidence may be provided on the issues involved.

CHAPTER V

SUMMARY

The purpose of this last chapter is not merely to summarize what has been reported in earlier chapters; it is more or less an attempt to tie the loose bits together. There is a review chapter by chapter, and a summary of the tentative conclusions reached, as well as the hypotheses suggested for further investigation. The chapter closes by outlining the ways in which the study may be considered a contribution to knowledge.

Chapter by Chapter Review

Chapter 1 introduces the problem in question form: how may achievement examinations be improved in use so as to better define and attain the objectives of classroom instruction? This problem is the conclusion of an analysis. Examinations perform various useful functions. Nevertheless there is some dissatisfaction, however slight, among the general public and among educators and students in particular. Banesh Hoffmann (op. cit.) would attribute this largely to the poor quality of multiple-choice test items. While not denying

Hoffmann's charges an opinion is expressed that in the case of classroom achievement examinations the chief cause of dissatisfaction, where it exists, may be the fact that some teachers tend to over-emphasize the assessment function. This is an opinion, to what extent it is tenable may be judged by surveying students' attitudes toward examinations, and in particular discovering their perceptions of the functions of examinations. The dissatisfaction and the assumed major cause beg the question stated earlier as the problem for this study.

It is suggested that the problem may be solved by emphasizing the learning function of classroom examinations. One way to achieve this end would be to require the student to take the examination twice, first in class, and second outside class; in addition he may be required to evaluate his performances before having any feedback from his teacher. The study is therefore concerned with testing the assumption and the suggested solution.

The suggested solution also rests on the literature evidence--that examinations are a learning device. The reader is therefore introduced to a few of the previous studies which arrived at the conclusion that examinations perform a learning function. Jersild found this so in all but one of his replicated experiments. Standlee's investigations of the contribution of quizzes to learning produced results which were clearly in the predicted direction though the statistical

tests revealed no significant difference among the treatment conditions. In a similar study by Fitch (et al.) it was concluded that frequent testing resulted in "superior achievement." Page (op. cit.) who carried out his studies in a secondary school setting was interested in the "knowledge of results" aspect of examinations, and in particular on the effect of "teachers comments" on one examination on performance in subsequent examinations. His conclusion is worth quoting again: "when the average secondary teacher takes the time and trouble to write comments . . . these apparently have a measurable and potent effect upon student effort . . . or whatever it is which causes learning to improve." Such is the evidence on which the suggested solution to the stated problem is based.

The purposes of this study may be re-stated in different words, these were:

1) to test some methods which were supposed to be likely to increase the learning value of examinations;

2) to survey and describe the attitudes towards examinations and grading of those particular students in which the methods were tested, and

3) to develop a scale battery for the purpose of carrying out the survey referred to in (2).

The hypotheses were that each of the experimental groups would excel the control group in achievement of course objectives and that one of the experimental groups--the one that received double treatment, would excel all other groups in the said achievement as measured by the final examinations. It was further hypothesized that the "experimental" groups

would show a more favorable attitude towards examinations and grading than do the Control group.

The second chapter deals with the methods of investigation. The three treatment conditions described correspond to the means suggested for emphasizing the learning function of examinations. The criterion measures were of two types--the course end examinations and an attitude scale specially developed for the study. Other chief topics described include the population and the sample used, and the general procedure. Subjects were sampled from two courses in the College of Education, Michigan State University. They carried out the treatment exercises by following appropriate instructions given to them by their instructors.

The results are presented in chapter III. To investigate the hypotheses of the study the treatment effect was first tested by the Friedman X_r^2. In experiment 1 the effect was significant at the hypothesized level of five percent. A second test of the same effect by the Covariance method was not significant. The means for T_1 and T_3 were in the predicted direction; so were the means for T_2 and T_1. But none of these were significantly different. However, the mean for T_2 was significantly larger than the one for T_3 at the hypothesized level of five percent, and the Covariance analysis showed a risk less than ten percent for rejecting the null hypothesis. No treatment effects were found in experiment 2. In both experiments the results on the attitude measures were not always consistent with the hypothesis.

Chapter 4 discusses these results. The Friedman X_r^2 test revealed a significant treatment effect on achievement while the t-test comparisons located the source in T_2. This applied to experiment 1 only. The discussion brings out some of the limitations that might have eclipsed the effects of treatment, if any, when T_1 and T_3 were compared in experiment 1, and when all groups were compared in experiment 2. The chief of these limitations was weakness of control: All groups had the opportunity to use the previous examinations as study guide. Another possible source was the fact that the treatment was misapplied in experiment 2 at the beginning. The chapter goes on to explain the modi operandi of the T_2 condition. It stimulated extra effort; it served as a study guide; it taught test-wiseness and developed an examination set; it provided a reinforcing practice session. Such conditions would tend to increase students' achievement; but such tendency was evident in experiment 1 only.

No definite conclusions could be made on the effect of treatment on achievement of course objectives when the two experiments were viewed together. However, there was in experiment 1 a weak evidence that the second treatment condition tended to increase learning and achievement. It must be emphasized that such conclusions are highly tentative, based, as they are on a weak evidence.

The results of the Attitude measures were not definitive. The chapter points out that this may be accounted for by the total University environment. The attributes of the attitude objects are not concentrated within one course. "If so,"

quote from that chapter, "it would require a longer period of treatment within a wider range of situations in the University for the treatment to work, assuming that it was potentially effective."

The observation is made, however, that there is some evidence of construct validity of the scale battery. It would be expected that students "high" on the learning and motivating function factors would be low, relatively, on the dysfunction and pressure-anxiety factors. This is precisely what the results indicated.

Questions are raised on behalf of the student. Does the treatment provide any benefits other than the learning function? It would appear it does, if applied in the manner prescribed. The self-evaluation exercise may help in the objective development of self-concept and potentially it may have a "mental hygiene" value which could be exploited in a classroom situation. But this is an opinion founded on meagre evidence, though testable in a properly controlled study. Could the time involved in carrying out the treatment not be more profitably used in extra reading and other assignments? Perhaps not; the "treatment" specifically asks the student to make use of "all resources, excluding instructors and fellow students." The student is guided, yet is left free to search independently and the gains derived may be as much as, if not more than what he would have had from extra assignment by the teacher, without the treatment. On the question of student load it is assumed that the majority of students will learn

to adjust their own load should such conditions as described here be adopted across the University.

The teacher critic may also have some legitimate problems. Time is limited for him to discuss discrepancies of self-evaluations with students; time is limited for him to write comments on students examination scripts. There is no dispute of the fact that much time is involved, especially where classes are large and where part-time teachers are in charge. However, this burden has to be weighed against the potential benefits of individualized instruction. The teacher is also concerned about the "loss" of test items. But if the test items are good valid items it seems doubtful that there is any "loss" in view of the teacher's professional role. In any case, he may be consoled by such research evidence as provided by Ebel: "old" items may be remodelled and re-used as valid test items even when they are exposed.

Table 6 shows that on the average students perceive grading more as a motivating factor than a learning function factor; and this pattern is consistent across treatment conditions. This is an interesting and useful finding for the group in view of the fact that most of them "hate" grading as that table also shows.

The last section of the chapter considers the need for further study. The evidence of this one does <u>not</u> establish a case, but it has demonstrated a trend worth investigating. But the would-be investigator is reminded of the limitations of this study: in particular the control was weak, and the

instructors were not sufficiently involved. With these and other limitations removed the study may be designed to last at least two terms (or semesters) and to include other departments in the University, and possibly other Universities. Furthermore, another criterion measure is suggested: this might be a measure of retention after one term (or semester) has elapsed following treatment.

Hitherto the review has focused on the main study. The subsidiary one on attitude scale development will now receive some attention. Its significance stems from the fact that it provided the measure for testing one of the main hypotheses. Obviously, if the scale that resulted is not valid little meaning, if any, can be attached to the section in which the scale is used.

The scale was assumed valid before it was put into use, and there were reasons for taking this position. As discussed in Appendix A, it is based on a defined concept of attitude. The multidimensional model provided is the outcome of a reasoned criticism of the traditional linear model. The steps by which the attitude statements were constructed followed closely the proposed model. According to Thurstone, attitude can be inferred from opinion statements. Such statements were obtained from a small sample of students and content analyzed. The categories devised were in line with specific attributes of the objects and the two dimensions of attitude defined in the model. The significant statements derived from the analysis formed the nuclei of statements which were rated by

a group of ten judges. Statements selected were fairly homogeneous within the four original scales. The responses obtained from a representative sample of students were factor analyzed. The factor output confirmed and clarified the original categories into which the statements were grouped. There was therefore a strong beginning evidence that the items were measuring some homogeneous trait, the limitations of "response-sets" notwithstanding. The statements selected had loadings of .40 or greater in the various "factors" revealed. It is relevant to observe that no assumption is made that the factor analysis results are enough to demonstrate the validity of the scale; but they are necessary in a program of construct validation of which this scale provides a type.

The theoretical model that forms the basis for the scale is hypothesis--generating, and this would be another direction in which the validity of this scale may be investigated. As mentioned earlier it may be hypothesized that individuals who score high on the learning function factors would score low on the pressure-anxiety factors. The results of the main study would tend to support such an hypothesis. In short, the methods by which the scale was developed, the factor analysis results and the pattern of responses when the scale was first put to use--all these are evidences that the scale is fairly valid.

Contributions to Knowledge

Now that the whole study has been reviewed it might be asked: what contributions does it make to the field of education? What contributions does it make to the general field of knowledge? This is a legitimate question; but it is the reader that must have the last say; he must judge whether this study, in parts or in whole is a contribution to knowledge: that is his prerogative.

Even so, the writer has as a duty to set down specifically what he considers to be the areas in which the study may be considered an extension of knowledge. The attitude scale comes first to mind. In these pages has been presented a scale battery to measure student's attitude towards such crucial issues in the educational curriculum as classroom achievement testing and grading. No educator will doubt the impact of these aspects of the curriculum in the learning situation. The argument pressed here is that attitudes are determined by the attributes of the psychological object. If, and when the validity of this scale is attested beyond all reasonable doubt the teacher may use it as a barometer to check the quality of his examinations and grading policies as reflected by the perceptions of his students. Such use will benefit education if it should by any way subscribe to the fulfillment of its objects.

In the general field of knowledge the theoretical model of attitude may be considered an extension. The existing

attitude scales expressly consider attitude as bipolar. The view taken here is that in the domain of attitude "good" may not be the opposite of "bad"; they may very well be on correlated dimensions: the relationship is inverse, but not perfectly so. Furthermore, the view expressed by other critics of the theory of attitude is accepted that attitude is multidimensional in character. The present study goes further to locate these dimensions in the attributes of the psychological objects, and these are incorporated into the theoretical model of attitude--incorporated, that is, in a way that is of practical value. As has been illustrated in the present scale the model provides a scheme for writing attitude statements and the resulting scale does spell out as it were areas in which the attitude object may be manipulated should one wish to effect attitude change. If, and when this approach is tried by other investigators and found to work, the model would then indicate its claim to be a contribution to knowledge.

There is nothing new if the main study merely harps upon two of the roles of the teacher-educator--as a promoter of learning and as an evaluator of learning. Apparently, these roles are conflicting. In fact, there is literature evidence (Fitch, op. cit.) suggesting that each is best served when divorced from the other. This study would reject such a suggestion. Instead, it argues that the apparent conflict may be resolved if the teacher-educator puts emphasis on the learning function of classroom examinations. Let us put aside

for the moment any question on how the emphasis is to be laid. The prescription: "emphasize the learning function of classroom achievement examinations" is brief, simple, and, one may add, easy to apply. Should the weak evidence provided by this study be collaborated and strengthened by other investigators such that educators will deliberately seek to derive this potential value of examinations, then it would be a little contribution to have drawn the attention of educators to a common sense and practical way of resolving an apparent conflict in their roles.

Some methods have also been provided to translate the said emphasis into practice. As a matter of fact, there is nothing new or original in the methods. But they are presented as a package deal, and in a way is a unique combination; the incorporation of students' self-evaluations deserves special mention in this respect. Above all, the study has provided an evidence in one case that the treatment would tend to increase achievement; other evidences are necessary to throw more light on the issues involved. Here again the methods may be considered a contribution to the practical problems that face the classroom teacher, if and only if, when tested in a variety of situations, the results should overwhelmingly point to the desired direction.

Summary of Tentative Conclusions

1. The repetition of examination performance in non "examination conditions" would tend to increase students'

achievement of course objectives, provided the examination is made up of good valid test items.

2. The requirement that students evaluate their examination performance before receiving the teachers' feedback would tend to increase their achievement of course objectives, provided the examination is made up of good valid items. These tentative conclusions are in essence hypotheses for investigation, since they are based on weak evidence. In any case they are made with particular reference to the type of situations as described in experiment 1.

3. It may be concluded from the results of the attitude survey that most students in the population studied perceive grading as performing some useful function: it motivates learning.

Summary of Testable Hypotheses

1. There is an inverse relationship between the attitude responses on the learning function factors on the one hand and on the dysfunction and pressure-anxiety factors on the other.

2. There is a direct relationship between students' perception of the learning function of examinations and their achievement of course objectives as measured by examinations.

3. If students are required to evaluate their examination performances before having their instructors' feedback then it would follow that:

a) their attitudes would tend to be more positive than negative towards examinations and grading,

b) they would develop a more positive than negative self-concept.

Summary of Conditional Contributions

1. The field of education is presented a scale for measuring students attitudes on such crucial aspects of the curriculum as examinations and grading.

2. The results of the scale may be used as basis for effecting attitude change in a desired direction--to promote students' learning.

3. The results of the scale may also be used by the teacher as a barometer to check the quality of his examination and grading policies.

4. Theoretically the multidimensional character of attitude is defined with reference to the anchoring attributes of the psychological object.

5. Theoretically the evidence provided by the maiden use of the scale does not support the traditional linear continuum and bipolar model of attitude.

6. The multidimensional model proposed may be applied by social scientists in the study of attitudes.

7. Lastly: the field of education is provided with a prescription which may resolve the apparent conflict in the roles of the teacher as a promoter and as an evaluator of learning. "Emphasize"--the prescription says--"emphasize the learning function of classroom achievement examinations."

It must be emphasized that these contributions are conditional: they are conditional on collaborating evidence from other investigators.

Conclusion

The problem of this study is stated in question form: How may achievement examinations be improved in use so as to better define and attain the objectives of classroom instruction? Tyler (op. cit.) is voicing the same problem in different words: "We who are concerned with the improvement of education and the effectiveness of learning must consider how to achieve the maximum good potential in testing and to minimize and eliminate the bad. . . ."

This study suggests some methods that may be applied to meet the need expressed in this quotation, and in the problem statement. There was some evidence in one of the experiments that the methods may lead to increase in students' achievement of course objectives. Probably there would also be effects on students attitudes if the treatment is widely applied, and is allowed sufficient long time--to work. However, all such conclusions will remain tentative until enough collaborating evidence is available. But suppose the wanted evidence turns out to be contrary, then the problem remains--unsolved, but not beyond solution.

LIST OF REFERENCES

Ebel, R. L. Measuring Educational Achievement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1965.

_____. Personal communications (1968).

Edwards, A. L. Techniques of Attitude Scale Construction. New York: Appleton-Century Crofts, 1957.

Fishbein, M. "Attitude and the Prediction of Behavior," a chapter in the book he edits, Attitude Theory and Measurement. New York: John Wiley and Sons, 1967.

Fitch, M. L., Drucker, A. J. and Norton, J. A. "Frequent Testing as a Motivating Factor in Large Lecture Classes," J. Educ. Psychol., 42, 1951, pp. 1-20.

Kerlinger, F. N. and Kaya, B. "The Construction and Factor Analytic Validation of Scales to Measure Attitudes Towards Education," Ed. and Psychol. Meas., 19, 1959, 13f.

Freeman, L. (Ed.) "The Measurement of Opinion and Attitude in Young," Handbook of Social Psychology.

Hoffman, B. The Tyranny of Testing. New York: Crowell-Collier Press, 1962.

Jersild, A. T. "Examination As an Aid to Learning," J. Educ. Psychol., 20, 1929, pp. 602-609.

Magnusson, D. Test Theory. Reading: Addison-Wesley Publishing Co., 1966.

Meyer, G. "The Effect on Recall and Recognition of the Examination Set in Classroom Situations," J. Educ. Psychol., 37, 1936, pp. 81-99.

MSU CISSR. Michigan State University, Technical Report No. 34 issued by the Computer Institute for Social Science Research. 1967.

Page, E. B. "Teacher Comments and Student Performance: A Seventy-Four Classroom Experiment in School Motivation," J. Educ. Psychol., 49, 1958, pp. 173-181.

Scheffé, H. The Analysis of Variance. New York: Wiley, 1959.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company, Inc., 1956.

Standlee, L. S. and Popham, W. J. "Quizzes' Contribution to Learning," J. Educ. Psychol., 51, 1960, pp. 322-325.

Stone, G. "The Training Function of Examinations: Retest Performance as a Function of the Amount of Critique Information." Research Report No. AFPTRC-TN-55-8. United States Air Force Personnel Training Research Center, Lackland Air Force Base, San Antonio, Texas, 1955.

Tyler, R. W. "What Testing Does to Teachers and Students," in Anne Anastasi (Ed.) Testing Problems in Perspective, American Council on Education, 1966.

Vernon, P. H. The Structure of Human Abilities. London: Methuen, 1961.

Warrington, W. G. Personal communications (1968).

Wrigley, C. Personal communications (1968).

General References

Aiken, L. R. and Dreger, R. M. "The Effect of Attitudes on Performance in Mathematics," J. Educ. Psychol, 52, 1961, pp. 19-24.

Bateman, R. M. "The Construction and Evaluation of a Scale to Measure Attitude Toward Any Educational Program," J. Educ. Res., 36, 1943, 502f.

Cronbach, L. J. "Further Evidence on Response Sets and Test Design," Educ. and Psychol. Meas., 10, 1950, pp. 3-31.

Green, B. F. "Attitude Measurement," in Lindzey, G. (Ed.) Handbook of Social Psychology, Cambridge, Mass.: Addison-Wesley, 1954.

Thurstone, L. L. The Measurement of Values. Chicago, Illinois: University of Chicago Press, 1959.

Wang, C. K. A. "Suggested Criteria for Writing Attitude Statements," J. Soc. Psychol., 3, 1932, 367f.

APPENDIX A

The Students Attitudes Towards Examinations and Grading Scale Battery (SATEG)

SECTION HEADINGS

ABSTRACT

The test battery proposed in this Appendix is an instrument for measuring students' attitudes towards examinations and grading.

Attitude itself has evolved from a unidimensional to a multidimensional concept. However authorities are not agreed on what its relevant dimensions are; nor has a clear attempt been made to reflect this multidimensional character in current Attitude Measures. Here attitude is defined as a predispositional set of like and dislike feelings towards a psychological object. Outwardly its dimensions are two--the like (positive) and the dislike (negative) feelings, which are not necessarily bipolar since they are anchored on different attributes of the psychological object. A model is presented in which the attitudinal disposition is depicted as a basal plane. From an origin on this plane emanate two separate vectors representing the positive and negative dimensions of attitude. The growth of these is determined by the attributes of the psychological objects. These attributes are called anchors, and serve to elucidate the multidimensional character of attitude. This is the theoretical basis against which the proposed battery is to be appreciated.

Examinations refer to all written forms of classroom achievement testing in which the results are used in making academic decisions on students. Grading on the other hand refers to a system of evaluation of academic performance in which some ranking procedures are used to reflect either the relative standing of a student as compared to his peers, or his achievement in a defined content area.

First, free reactions of a small number of students on these objects were obtained through the administration of an open-ended questionnaire. The returns were content analyzed, and significant statements extracted to form the nuclei of attitude statements. These were written to focus on specified attributes of the objects, and classified under two main directions--positive and negative. Ten raters then judged and Q-sorted the statements on an eleven-point scale. The final statements were selected on the basis of their median values, with preference for high extreme values only. Such statements would tend to be homogeneous and would discriminate satisfactorily. The selected items were grouped under four scales: examination-positive (EP), examination-negative (EN), grading-positive (GP), and grading negative (GN). Finally they were administered to a sample of 585 students. Subjects responded on a specially defined Likert-type scale. The results were factor analyzed.

In line with the theoretical model six factors were hypothesized. But the analysis produced eight Varimax

factors. Loadings on these are the only criteria used in the final selection of items to comprise the battery. The bulk of the test statistics and details of the methods described are presented in the sub-appendices which form an integral part of this paper. The body text has some of the chief statistics: for example the K-R_{20} reliability coefficients are reported as .798, .791,.812 and .746 for each of the four original scales respectively.

An illustration is also provided for interpreting the scores as "low" or "high" on the reference factors. This is based largely on the mean item responses for each factor sub-scale. A profile of the sample's attitudes is also presented. Generally students are "low" on the positive learning function factors, and relatively "high" on the pressure-anxiety factors. This fact would suggest a general hypothesis that the higher a student is on the learning function factors the lower he would be on the pressure-anxiety factor and consequently the higher the amount and quality of learning he would attain. The testing of such an hypothesis may be incorporated into a program of construct validation of this scale battery.

This study provides evidence which tends to support the model and the way it defines the multidimensional character of Attitude. An instrument like this may be used for attitude survey purposes. It is also of diagnostic value and suggestive of ways of effecting attitude change.

However the battery would remain a research instrument--till further evidence is available from other investigators.

GENERAL INTRODUCTION

The concept of attitude has undergone some evolution. Formerly it was unidimensional; now this view will not satisfy all. There are some authorities who consider attitude as made up of several components, including the affective, the cognitive and the conative. Fishbein (1967) does not favor this type of approach however; rather he would emphasize the affective aspect as the essence of attitude and the cognitive and conative as its "determinants" or "consequences." Thus like Edwards (1957) he would prefer Thurstone's notion of attitude as "the degree of positive or negative affect associated with some psychological object." The writer is inclined to accept the affective aspect as the essence of attitude; but he does not find the unidimensional approach very satisfactory if this means that attitude is to be represented in some linear continuum. A more realistic picture would be that of a predispositional base of like-dislike feelings towards an "object." From such a base or "set of reaction tendency" (to quote Freeman, 1957) emanate what may be described as "vectors" symbolizing different directions or dimensions of the attitude. Some of these directed vectors are towards and favorable while others

may be away from and unfavorable to the object. The length of each vector would symbolize the strength of the respective dimension of attitude. It would be convenient for measurement to focus attention on the two most significant attitude vectors one is directed towards and the other away from the object; they may be described as "positive" and "negative" respectively, provided no assumption is made that they are diametrically opposed. These two vectors in turn each have sets of subsidiary branching--vectors which symbolize the attributes of the psychological object.

These branching vectors in effect determine the lengths and strengths and hence the dominance of the respective dimensions whose existences they maintain. In this sense they may be referred to as the ultimate anchors of the attitude. These may well be the foci of attitude statements.

The use of "vectors" here is somewhat loose, but would serve for the purpose of analogy. Both the "positive" and "negative" vectors are in the same (attitude) "space"; the attribute vectors may be in a different (object) "space", but have direct links with the former, and both converge upon the attitudinal base. Perhaps a pictorial account may help to clarify the model. Figure 1, on the following page, serves this purpose.

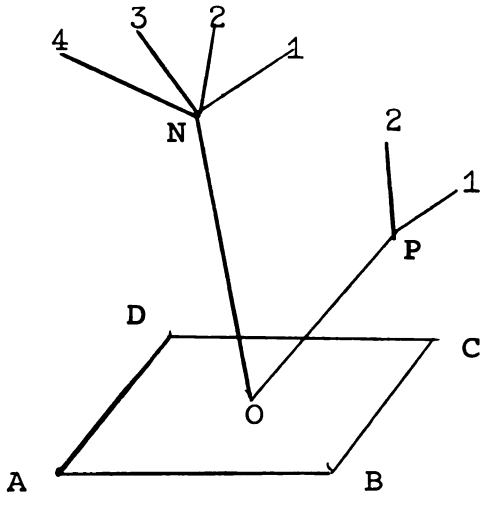


Figure 1. The Attitude Model.

In the above diagram ABCD represents the predispositional base of like-dislike feelings, with O as its outgrowth point. OP and ON represent the "positive" and "negative" dimensions respectively: their slopes reflect their nature. The attribute vectors are P_1, P_2 as well as N_1, N_2. They sustain or serve as anchors to the "dimension" vectors--that is to say: the attribute of an object accounts for the direction or "dimension" of attitude developed about the object. It is to be observed that the lengths of the vectors vary; moreover the number of the attribute vectors on each dimension also vary. Furthermore it is conceivable that for certain objects the attributes may support growth along one dimension only. In such a case the attitude may be described as all-out "positive" or all out "negative" as the case may be. The traditional model makes no provision for these attitude anchors; besides the dimension vectors are made to

collapse on the base, end to end thus producing a bipolar continuum. The view taken here is that such a model is an oversimplification.

The present model may be justified. Given a psychological object an individual's attitude in most cases is not an all out feeling of like or dislike. It may well be a mixture of the two. Certain attributes of the object may stimulate and sustain like feelings while others induce dislike feelings. In other words while an individual may profess a favorable attitude towards an object he may be found to have some unfavorable attitude also. This is no inconsistency, but the hard fact of human experience.

It is necessary to reiterate what has been said of the proposed multidimensional model. The scale battery to be presented cannot be fully appreciated without this model. For this study attitude will be defined as a predispositional set of like and dislike feelings. Two significant vectors emanating from this set may be described as "positive" and "negative". But these are themselves sustained by the attributes of the psychological object. Measurement of attitude would therefore be concerned with the problem of placing individuals on the attribute vectors. A description of a person's attitude on a profile of such significant "vectors" would be nearer to reality than the one on the traditional linear continuum.

The psychological objects of interest also deserve some comment. Examinations refer to classroom achievement testing

involving the administration of quizzes, mid-terms, and finals, made up of objective or essay items--provided that students performances form the basis on which academic decisions are made, viz Pass/Fail, Credit/No Credit or the award of grades. Grading refers to evaluation of student's academic performance in examinations and/or other aspects of the curriculum by using letters (e.g., A, B, . . . F) or numbers (e.g., 4:5 4.0 . . . 0.0) to classify students according to their achievement relative either to their peers or to a defined content area, provided the grading system also involves the report of "grade-point averages".

As the objects are defined above it is evident that the reference population to which the proposed scale battery may be applied consists of "students". In its initial development University students were used, but it is hoped that the language adopted in the final form is such as to make it applicable to High School students, either directly or with minor alterations. Thus, the purpose of this scale battery is to ascertain the predispositional set of like-dislike feelings of students towards examinations and grading, to map, as it were in a profile, the relative strengths of these like-dislike feelings. The ultimate aim is to provide a means for a fairly accurate description of the attitude.

It may be asked: "What is the use of knowing student's attitudes towards examinations and grading?" "What can one 'predict' by having such information?" Fishbein (op. cit.) feels that "the most important determinants of behavior may

be variables other than attitude." This may be so with most psychological objects. But in the learning situation the attitude of the learner may exert quite a significant effect, if not the most important effect on the learning outcome. It may be that as yet psychologists have not been able to develop a tool to identify the effect of attitudes on learning. Provision of a valid scale to discern the attitudes on crucial issues in the curriculum may be a necessary first step. One would venture to hypothesize that students' attitudes on issues like examinations and grading partly determine the amount and quality of learning attained. It may also be argued that in a student or child-centered system of education the attitude of the student or the child should not be ignored. Furthermore, findings from an instrument which specifies the attributes of the attitude object will suggest which aspects of the object to manipulate, as it were, should one wish to effect an attitude change. This brief discussion of the predictive and other uses of this instrument is part and parcel of the overall purpose for which it is designed.

The plan followed in the development of this Scale Battery may be listed:

1. Administration of an open-ended questionnaire to a small sample of students
2. Content analysis of responses obtained in step 1 above.
3. Development of attitude statements from the results of step 2 (above).

4. Determination of statement values by the use of judges employing Q-sort and Rating techniques, and selection of statements with values at or above the median

5. Preliminary try-out of the selected statements

6. The main try-out, using a sample of 585 students

7. Factor Analysis--to ascertain the factorial validity of the scales and use factor loadings for item selection.

These stages are discussed in full in the pages which follow.

THE OPEN-ENDED QUESTIONNAIRE

The plan was to administer an open-ended questionnaire to a small sample of students. There was no attempt to use any controlled sample since the objective was not to make generalizations from the returns. Considerations of the immediate use for which the intended scale was to be applied necessitated a preference for students in Education. This is evident from the tabular illustration below where the respondents are broken down according to the courses in which they were enrolled.

Course	Number of Students
Ed. 200	11
Ed. 450	6
Ed. 867	16
Phys. 827	1
Mth 433	3
Mth 215	9
Psy. 312	1
CEM 311	3
TOTAL	50

It must be emphasized that the purpose behind this step was to elicit the attitude in question, and to check whether in fact students may be broadly categorized into those who favor and those who do not favor examinations or grading. The typical language of each group would then be ascertained and used as a basis for developing attitude statements.

Fifteen items comprised the Questionnaire. Item 4 reads: "What reactions have you had to the examinations you have taken in your College and University experience?" The other items and the exact format of the Questionnaire may be found in Sub-appendix (a).

CONTENT ANALYSIS

There was one and only one purpose for the content analysis of the free response in the returned questionnaire: to ascertain the typical language of students with positive attitude and those with negative attitude towards the psychological objects investigated. The search was therefore for significant affective words, phrases, clauses and sentences in the unit of analysis, which in this case was the whole response to each item. These "significant" words, etc., were categorized according to the scheme to be illustrated presently.

The setting up of content categories posed some problems. The guiding principle was the theoretical model described as the basis for the intended scale. Attitudes have dimensions

maintained by the attributes of the psychological object. The content categories should in turn reflect the attributes of the object. Following this reasoning nine categories were arrived at as follows:

1. Perceived function/meaning of object
2. Statement on efficiency or inefficiency
3. Expressed statement of preferences
4. Expressed opinion with emotional overtones
5. Expressed opinion with very intense emotion
6. Indication of satisfaction
7. Indication of dissatisfaction
8. Suggestions of alternatives
9. Miscellaneous (unclassified) reasons stated.

A scheme for coding and general procedural steps in the analysis were prepared. Two questionnaire copies were then content analyzed. It was feared that the content categories lacked the qualities of objectivity, reliability and validity. The fears were confirmed when another analyst[1] was engaged. Discussions that followed led to the reduction of the number of categories to four. These were:

1. Statement of function (e.g., feed-back)
2. Statement of preferences
3. Statement expressing or implying emotion, and
4. Statement offering suggestions.

---

[1]The writer is very grateful to Ogunniyi Omotosho for the role he played as analyst in this aspect of the study. "Tosho" is currently finishing his Ph.D. dissertation.

Based on these reduced categories and on improved instructions another questionnaire was analyzed independently by the writer and the engaged student. Agreement was perfect on every item. The rest of the scripts were then analyzed by the writer only following the improved scheme, which is given in full in the sub-appendix (b).

The words or phrases selected as "significant" tended to the extreme: for example, "examinations should be completely abolished." It was thought that extreme statements would discriminate better than moderately affective ones; moreover they would help to make the scales homogeneous.

ATTITUDE STATEMENTS

The content analysis exercise revealed that two categories were the richest and most appropriate as sources for attitude statements. These were (1) Statement of function, and (3) statement expressing or implying emotion. The first clearly focuses on one dimension of the psychological objects while the second touches on varied aspects, including for examples, administration procedures of both examinations and grading, types of the examination or grading system and quality of test items. The "significant" statements selected under the two categories were then tabulated and the frequency of each statement across the fifty respondents was determined. The following extract from the work-sheet illustrates this point with respect to grading:

Frequency of Significant Statements

Statement	Undergrad Education Courses	Graduate Education Courses	All Other Courses	Total
"motivates the student"	4	4	2	10
"abolish"	5	4	6	15

A majority of the significant statements selected were the most frequent. However judgment was exercised to include those not necessarily the most frequent but thought to be referring to important attributes of the psychological objects. Finally sentences were constructed using the significant statements as nuclei. The sixty-five initial attitude statements are shown in Sub-appendix (d).

STATEMENT VALUES

The next step was aimed primarily at assigning numerical values to the statements. The secondary aim was to use other people to judge whether the statements were meaningful, clear and unambiguous. The original plan was to use two groups of advanced graduate students for this exercise. One group would "judge" and "Q-sort" the statements while the other would "judge" and "rate" them. In judging the subject had to say whether the statement reflected a positive favorable attitude or an unfavorable negative attitude towards the object. Rating involved weighing each statement singly and assigning a value to it. In Q-sorting all the statements

within a group were viewed together and assigned relative values, so as to produce a near-normal or rectangular distribution. The eleven-point scale was to be used in both cases. Full details of the instructions may be found in Sub-appendix (c).

In a pre-session it was found that most of the raters assigned extreme values to the statements. There was very little discrimination. The use of this technique was therefore discarded. Incidentally it should be mentioned that the results from the two techniques were to be compared and averaged. As it turned out, such averages would have been meaningless.

A forced-choice "Q-sort" technique was adopted because discrimination among the items was possible. Ten judges[1] were engaged. The value of each statement was the median of the values assigned by the ten. The extract from the worksheet, on the following page, illustrates the procedure.

In an eleven-point scale the median value is six. The criterion for selecting an item was therefore a minimum value of <u>six</u>. There was however another requirement that at least

---

[1]The writer is grateful to the persons listed below for the role they played in this aspect of the study. The first two hold Doctorate degrees and are Assistant professors in the Departments of Education and Psychology respectively. Of the rest six are advanced graduate students working for their Ph.D. degree, and two are Master's candidates (M). Dr. D. Freeman (Education), Dr. A. M. Barclay (Psychology), Miss Gisila Dieze (M), Miss Jody Anderson, Terry Almquist, William E. Martin, William S. Beavers, Paul David Goff (M), John Hoogstra, and Dick Bate.

Reference Object, Direction and Value of Statements

Judges			S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	Total	Median
Item No.	Obj.	Dir. and Val.												
3	Exam	Pos.	x	x	x	x	x	x	x	x	x	x	10[a]	
		Neg.												
		Val.	11	9	7	5	11	6	3	8	11	11	b	9[c]
	Grad-ing	Pos.												
		Neg.												
		Val.												

[a]This total reflects the absence or presence of ambiguity in the statement. In this case all judges agree the statement is referring to examination, and expresses an attitude that is positive, i.e., favorable towards it.

[b]This total is irrelevant and so was not calculated.

[c]This becomes the value of the statement (the median values for the other statements are given in Sub-appendix (d).

fifty-five items should be selected from the original sixty-five. To meet this requirement also five items were selected, each with a value of five. There was no absolute need to calculate and use indices of dispersion since the aim was not to produce a scale purely on the Thurstone model.

It is necessary to explain the purpose of statement "values" at these early stages in the development of the scale battery. The values were not to be used in the Thurstone style: this must be emphasized. They were to be used as aids to developing a homogeneous scale. Suppose for example, that a statement has been judged to be of "negative" direction; if it is further assigned the value of one then it represents a statement that is tending towards a positive direction; if on the other hand it is assigned the value of eleven it can be safely assumed to represent an extremely negative statement. Ideally only items, each with value eleven would be selected since as stated earlier there was reason to prefer high extreme statements. To attain this ideal is not impossible; all it involves is increasing the original pool to at least 500 items, carefully written with the same goal in mind. Actually the median value of the selected items turned out to be seven, and there were two items with value ten and another two with value eleven each. This is admittedly a poor approximation to the ideal, but was accepted as fairly satisfactory in the present circumstances.

Another comment is in place. Each statement was assigned a relative value within its own group. The direction of the attitude implied in the statement was not taken into account in assigning values. Thus, these values are not to be confused with the five-point-Likert-scale used in the final scale battery, as will be shown presently. In fact the exercise thus described is an elaborated example of stimuli scaling--the attitude statements are scaled; on the other hand the Likert technique scales persons. The two methods were therefore combined in the present study.

PRELIMINARY TRY-OUT

The preliminary try-out was necessary to check on the suitability of the format in which the battery is to be presented, to check also on the clarity of instructions, and again on the quality of the items. Moreover it would provide an opportunity to test the scoring procedure before a full scale try-out was launched. This last need emerged from discussions with Warrington (1968).

In view of the purposes just stated the "sample"--if it could be called one at all--was confined to three advanced graduate students[1] invited to respond to the items in their role as University students. Later they were expected to provide and did provide written comments as they felt necessary.

---

[1]The writer is grateful to the advanced graduate students named below for the role they played in this part of the study: Jack Hruska, W. Russel Harris, Glenn L. Starner.

As mentioned in the last section fifty-five items were selected. Respondents were expected to show the degree of their agreement/disagreement with the statements by assigning values using a five-point Likert scale. The points were defined as follow:[1]

1. No agreement whatsoever
2. Disagreement most of the time; agreement at few occasions.
3. Opinion hovers between agreement and disagreement equally.
4. Agreement most of the time; disagreement at few occasions.
5. Complete agreement.

There were four groups of items: Examination-Positive, Examination-Negative, Grading-Positive and Grading-Negative. From now on these will be referred to as EP, EN, GP and GN respectively. They are the four scales which constitute the scale battery. To simplify notations further they will also be referred to as Scales 1, 2, 3, and 4 respectively. No systematic order was employed in arranging the items in each scale; but the scales were chosen alternately, and no more than six items in the same scale were presented successively. The results from this investigation were as shown in tabular form on the following page.

---

[1]These definitions may be cumbrous; but the aim is to avoid the stereotype and thus hopefully minimixe response sets.

ATTITUDE SCORES PRELIMINARY TRY-OUT

	Scales			
Possible score	EP	EN	GP	GN
Possible score range	11-55	12-60	14-70	16-80
Cutting score*	33	39	42	48
Respondents				
S_1	17	32	33	57
S_2	37	25	46	35
S_3	14	46	17	59

*These scores are determined from the Likert point value of 3 as defined above. Respondents with scores above the cell entries here can generally be classified as being "high" on the attitude measured by the scale. The values vary with the number of items in each scale; no final selection of items was made, as yet.

RANK-DIFFERENCE CORRELATION COEFFICIENT*

	EP	EN	GP	GN
EP		-1.00	+1.00	-1.00
EN			-1.00	+1.00
GP				-1.00
GN				

*The high values are certainly an artifact of the sample size; does this also apply to the direction?

The preceding pattern of scores and the direction of the coefficients would be expected from the theoretical model; the absolute values were of no significance. This part of the exercise was therefore very valuable in that it also led to the improvement in the diction of some of the items and in the format of the instructions--all based on the comments from the respondents and other consultants. Of the fifty-five items used, forty-eight were retained--twelve each for the four scales EP, EN, GP and GN.

THE MAIN TRY-OUT

Two considerations determined the characteristics of the sample drawn for the main try-out phase. The first was the immediate population for which the Scale is designed. The Scale is directly applicable to a population of college and university students. It is assumed that the students of Michigan State University form such a typical population. The sample was drawn in such a way that the main departments of the University are represented. However, it was not random; judgment was exercised to make the selection include "juniors" and graduate students as shown in Sub-appendix (e).

The second consideration was the intention to factor analyze the returns--in an effort to test the validity of the theoretical model conceived as the basis for the scale battery. Accordingly, the size of the sample was planned at 600 at least. As Sub-appendix (e) shows, the actual returns

were 585 (incidentally twelve data cards were destroyed in process so that the final output involved 573 observations).

The questionnaire was administered by the instructors[1] responsible for the classes selected. Subjects responded to all items on a five-option IBM answer sheet. About fifteen minutes were sufficient to respond to all items. The scoring was done by the Office of Evaluation Services.

THE FACTOR EXPERIMENT

Both the theoretical basis for the battery and the hypotheses that may be deduced from the model may sound a little radical. It is therefore necessary to put them through a somewhat rigorous test as may be provided by factor analysis. In the first place the view is expressed that like and dislike attitudinal feelings are not necessarily on a linear continuum. Accordingly it was hypothesized that EP and EN scales represent two distinguishable "factors" and not one bipolar factor. Similarly GP and GN scales also represent separate factors. The model also depicts attributes of the psychological object as the anchors for attitudinal feelings. It would follow therefore that where a number of attitude statements focus on a well defined attribute of the object

---

[1]Space forbids the listing of the twenty-and-two professors who were not only willing to permit the use of their classes but also agreed to administer the questionnaire to their students in an effort to help keep "the experimenter out of the scene." The writer is deeply grateful to these professors and their students for their cooperation.

factor analysis would bring out a "factor" symbolizing such attribute. In the present battery development it was possible to focus a number of statements on the functions of the objects of interest. The content analysis exercise provided for this catetory. The second hypothesis was therefore that a "functional factor" would emerge from the analysis.

As mentioned above one of the richest content categories on which the attitude statements were based was the one in which emotion was expressed on diverse aspects of the objects. It was therefore not possible to formulate a well defined hypothesis in this area. At best it was hypothesized that a general attitudinal factor would also emerge.

The six types of factors discussed were clearly anticipated. But perhaps there might be another factor or factors engulfed in the general factor. With such reasoning the raw data was submitted for analysis in the hope that there would emerge "at least five factors".

The Rotation Techniques

The analytic procedures were repeated three times. In the first and second, half the observations were used--randomly divided; the third repeat involved all the observations. The Kiel-Wrigley criterion (MSU CISSR, 1967) was used in the rotation of factors for the two half samples, but the full sample data was rotated to ten factors.

Both the Quartimax and the Varimax methods of rotation were applied. Extracts from the final outputs are given in Sub-appendix (f). Only the loadings with value 0.40 or greater are shown on that table. The lower values may not be significant. The sample was split so that the factor patterns may be compared. Such comparison would throw light on the stability of the factors.

The full data analysis resulted in six Quartimax factors each of which has loadings on at least three variables. Three of these factors each account for at least five percent of the common variance. The other four factors may not be significant. The corresponding distribution for the Varimax factors is as follows: nine factors--with at least three variables, five factors, each accounting for at least five percent of the common variance and only one factor that may not be significant. Following Wrigley's (1968) suggestion the Varimax factors are adopted as the more appropriate in the present case. In fact there are also evidences in the literature (e.g., Vernon 1959, Kerlinger and Kaya, 1959) to justify this preference. But it is worth observing that both techniques of rotation produce more factors than were hypothesized. If the traditional model applied in this case there would have been at most three factors. Furthermore, the patterns across the three samples though not in perfect agreement are sufficiently similar, and tend to show the factors in the third analysis are stable. A full comparison of the Varimax factors across the three samples and the four Scales is

The Naming of the Varimax Factors in the Full Data Analysis

Factor 1: (16.08% of Common Variance) "(General) Learning Function"

Var. No.	Quest. No.	Attitude Statement	Loading
1	EP	Sum of scores on 12 items comprising Exam Positive Scale.	0.7262
6	EP_3	Of all teaching devices, examinations provide the most useful feedback.	0.4008
27	EP_{24}	Examinations provide the most satisfactory means for assessing learning.	0.5654
28	EP_{25}	Examinations are an indispensable feature of the University curriculum.	0.7194
29	EP_{26}	Without examinations, academic standards fall.	0.7911
36	EP_{33}	The discipline of examinations is vital to learning.	0.6628
38	EP_{35}	Abolition of examinations will in the long run lead to chaos in graduate education.	0.6348
3	GP	Sum of scores on 12 items comprising Grading Positive Scale	0.7071
16	GP_{13}	Grades provide a necessary incentive to hard work.	0.5269
23	GP_{20}	The grading system should be an integral part of the curriculum in higher education.	0.5212
24	GP_{21}	For the student, grades are a desirable aid to self-evaluation.	0.5033
25	GP_{22}	Abolition of grading would jeopardize learning at the University level.	0.7572
26	GP_{23}	Grading is a necessity if standards have to be maintained in University education.	0.7726
39	GP_{36}	I would campaign vigorously against any attempt to abolish grading at the University level.	0.5912
40	GP_{37}	Without grading the motivational function of examinations would be impaired.	0.5339
21	EN_{18}	Examinations should be abolished at the University level.	-0.4138
31	GN_{28}	Grading should be abolished at the University level.	-0.5152

given in Sub-appendix (g). The conclusion from that table is that the stability of the factors is not in doubt.

Seventeen variables have "significant" loadings on this factor; of these there are seven each belonging to the original EP and GP scales, and one each to the EN and GN scales.

On the positive side the theme is that both examinations and grading are relevant in the curriculum; the negative side is also clear: these aspects of the curriculum are not relevant and "should be abolished".

This factor shows up as bipolar, but very few negative items load on it and these negative loadings may reflect the particular wordings in variables 21 and 31. Perhaps a bipolar attitudinal factor may be an artifact of the language used in the statement. This will therefore be called the General Learning Function Factor. Future revisions will discard variables 21 and 31 and all such types.

Factor 2: (5.43% of Common Variance) "Examination Type"

Var. No.	Quest. No.	Attitude Statement/Description	Loading
22	EN	Sum of scores on 12 items comprising Examination Negative Scale.	0.5425
18	EN_{15}	Objective examinations are nothing more than a guessing game.	0.7024
44	EN_{41}	Examinations are nothing more than trickery.	0.6649

Apart from the EN scale only two other variables load significantly on this factor. One of them suggests this may

be an "Examination-Type" Factor. Further studies may investigate whether there is any such factor. It is worth noting that no items on grading load significantly on this factor. It is therefore peculiar to examinations, and provides another evidence that negative attitude may be on a distinct attribute of the psychological object.

Factor 3: (7.06% of Common Variance) "Pressure-Anxiety"

Var. No.	Quest. No.	Attitude Statement/Description	Loading
2	EN	Sum of scores on 12 items comprising Examination Negative Scale.	0.4576
19	EN_{16}	Examinations provide the student a frustrating experience.	0.5790
20	EN_{17}	I resent the pressure which examinations bring on me.	0.7110
43	EN_{40}	Examinations generate too much anxiety	0.7808
30	GN_{27}	Grades induce too much worry.	0.7459

Here again the only items that load significantly on this factor belong to the negative EN and GN scales. All the items provide "pressure" or "worry" or "anxiety" stimuli. This will therefore be called the "pressure-anxiety" factor. Examinations and grading go together, once again suggesting some common frame of mind, or reflecting the fact that the attitude dimensions and the supporting attributes are the same for both objects.

Factor 4: (7.69% of Common Variance) "Grade-Measure"

Var. No.	Quest. No.	Attitude Statement/Description	Loading
3	GP	Sum of scores on 12 items comprising Grading-Positive Scale	0.6159
14	GP_{11}	Grades are very effective for indicating students achievements of the course objectives.	0.6593
15	GP_{12}	Grades are a good estimate of the quality of learning that has taken place.	0.6262
17	GP_{14}	Given the word "meaningful" as indicating your opinion of grading, rate it according to the strength of this opinion.	0.5815
23	GP_{20}	The grading system should be an integral part of the curriculum in higher education.	0.4292
24	GP_{21}	For the student, grades are a desirable aid to self-evaluation.	0.4175
41	GP_{37}	The finer the grading system, the better it reflects the students' competence level.	0.5125
42	GP_{38}	Given the word "relevant" as indicating your opinion of grades, rate it to show the strength of this opinion.	0.5499
33	GN_{30}	Grades are no indication of what the student has learned in a course.	0.4467

With the exceptions of variables 23 and 24 (which also load high on factor 1) these items focus on the effectiveness of grading as a measuring instrument. That variable 33 loads with an opposite sign may be just an artifact of its wording ("no indication") and not necessarily that the factor is bi-polar.

This shall be called the Grade-Measure Factor. It is hard to explain why a similar item on examinations does not load high on this factor. Are the perceptions of these objects as measuring tools on different dimensions?

Factor 5: (6.54% of Common Variance)			"Hate"
Var. No.	Quest. No.	Attitude Statement/Description	Loading
4	GN	Sum of scores of 12 items comprising Grading-Negative Scale.	0.5703
34	GN_{31}	Given the word "evil" as reflecting your opinion of grading, rate it to show the strength of this opinion.	0.4535
49	GN_{46}	I have nothing for grades but pure hate.	0.6439
50	GN_{47}	Whoever put more grades into the scale should be hanged.	0.7378
51	GN_{48}	It is grossly unfair to award a graduate student a "D" or an equivalent grade.	0.5969
47	EN_{44}	Given the phrase ("a farce" as indicating your opinion of examinations rate it to show the strength of this opinion.	0.4394
48	EN_{45}	In my experience as a university student, examinations are the instructors' make-shift without any real value.	0.4330

Here as in Factor 1 the attitudinal disposition is the same for examinations and grading. That this is a distinct factor is further evidence that a negative attitudinal disposition may exist on a separate dimension.

This is named the Hate Factor; it is somewhat general in that the determinants of the "Hate" are not specified.

Factor 6:(3.42 of Common Variance)

Var. No.	Quest. No.	Attitude Statement	Loading
11	EP_8	Examinations make me feel happy and confident.	0.5787
35	EP_{32}	Examinations should be given more emphasis in the University curriculum.	0.4249
37	EP_{34}	Examinations make study exciting.	0.6315

This may be a general satisfaction factor--in opposition to the Pressure-Anxiety factor. Perhaps if similar items were included on grading they would also load on this factor.

Factor 7: (3.95 of Common Variance)

Var. No.	Quest. No.	Attitude Statement	Loading
46	EN_{34}	The examination system is entirely lacking in precision.	0.5209
47	EN_{44}	Given the phrase "a farce" as indicating your opinion of examinations, rate it to show the strength of this opinion.	0.4347
51	GN_{48}	It is grossly unfair to award a graduate student a "D" or an equivalent grade.	0.5351

It is difficult to explain why these items should comprise a separate factor. The last two also load significantly

on the "Hate" factor. It may not be a stable factor. Further investigations may reveal the nature of this factor, if at all it exists on a separate dimension. Meanwhile it will be ignored.

Factor 8: (4.91% of Common Variance) "Motivating Function

Var. No.	Quest. No.	Attitude Statement	Loading
5	EP_2	Examinations are the best means for motivating students to learn.	0.5770
7	EP_4	Examinations enforce my desire to learn.	0.5960
8	EP_5	Given the word "favorable" as referring to your feeling about examinations, rate it to indicate the degree of this feeling.	0.5331
16	GP_{13}	Grades provide a necessary incentive to hard work.	0.3984

The central thought in the first three items is that examinations are perceived to motivate learning. The loading of the last item on grading is below the criterion value of .40; however it is so close as to justify its inclusion here. This shall be called the "Motivating-Function" factor.

The statements which load on factor 9 (see the following page) seem to say that the psychological objects are worthless, or that they perform some undesirable function. This will therefore be called the Dysfunction Factor in opposition to the relevant Function Factors 1 and 4.

Factor 9: (4.52 of Common Variance)

Var. No.	Quest. No.	Attitude Statement/Description	Loading
12	EN_9	There is very little of instructional value in the content of examinations.	0.6338
13	EN_{10}	Examinations are redundant in the educational process at the University level.	0.6806
10	GN_7	Grading encourages students to cheat in examinations.	0.4539

It is worth observing that these variables do not load significantly on the first factor. There their loadings are 0.0806, 0.2254 and -0.0022 respectively. In other words the evidence is not very strong that either Factor 1 or Factor 9 is bipolar.

Generally the hypotheses were confirmed. Most of the "positive" statements came out under separate and identifiable factors; and so did the negative statements. Furthermore their identities have references or anchors in the attributes of the attitude objects. These attributes are reflected in the factor names suggested. However only limited success was achieved in separating the examination from the grading factors. Perhaps there is a natural linkage between them. It may also be that attitude factors are similar and parallel as shown in Figure 3, page 121.

SELECTION OF ITEMS AND PRESENTATION OF THE BATTERY

The table on the following page, shows the scheme used in making a selection of eight items for each of the four scales. The numbers appended refer to the items with the highest loadings on the respective factors. The table serves to emphasize the aims of the present battery. If attitude statements are anchored on well-defined attributes of the psychological object separate "factors" will emerge to symbolize these attributes. Furthermore the general nature of the attribute determines the direction of attitude, that is whether it is "positive" or "negative"--for or against. It may be added that this table also provides a scheme for writing new items. Ideally only unidimensional factors would serve in this scheme--to agree with the theoretical model, but factors 1 and 4 fail to meet this ideal.

The battery in the final form is reproduced on pages 109 and 110. Where groups of items belong to one factor they are arranged in descending order of the magnitude of their loadings, which were given earlier.

SCHEME FOR ITEM SELECTION

Factor	EP	EN	GP	GN
1 (Learning-Function Factor)	EP_{26};EP_{25} EP_{33};EP_{35} EP_{24}		GP_{23};GP_{22} GP_{36}	GN_{28}
2 (Examination-Type Factor)		EN_{15}		
3 (Pressure-Anxiety Factor)		EN_{40};EN_{17} EN_{16}		GN_{27}
4 (Grade Measuring-Function Factor)			GP_{11};GP_{12} GP_{14};GP_{28}	GN_{30}
5 (Hate Factor)		EN_{44};EN_{45}		GN_{47};GN_{46} GN_{48};GN_{31}
6 (Examination-Satisfaction Factor)	EP_{34}			
8 (Motivating Function Factor)	EP_{4};EP_{2}		GP_{13}	
9 (Dysfunction-Factor)		EN_{10};EN_{9}		GN_{7}

Scale	Factor	Item
EP	Learning-Function	1. Without examinations, academic standards would fall.
		2. Examinations are an indispensable feature of the University curriculum.
		3. The discipline of examinations is vital to learning.
		4. Abolition of examinations will in the long run lead to chaos in graduate education.
		5. Examinations provide the most satisfactory means for assessing learning.
	Exam-Satisfaction	6. Examinations make study exciting.
	Motivating-Function	7. Examinations enforce my desire to learn.
		8. Examinations are the best means for motivating students to learn.
EN	Exam-Type	1. Objective examinations are nothing more than a guessing game.
	Pressure-Anxiety	2. Examinations generate too much anxiety.
		3. I resent the pressure which examinations bring on me.
		4. Examinations provide the student a frustrating experience.
	Hate	5. Given the phrase "a farce" as indicating your opinion of examinations rate it to show the strength of this opinion.
		6. In my experience as a University student, examinations are the instructors' makeshift
	Dysfunction	7. Examinations are redundant in the educational process at the University level.
		8. There is very little of instructional value in the content of examinations.

GP	Learning-Function	1. Grading is a necessity if standards have to be maintained in University education.
		2. Abolition of grading would jeopardize learning at the University level.
		3. I would campaign vigorously against any attempt to abolish grading at the University level.
	Measuring-Function	4. Grades are very effective for indicating students' achievements of course objectives.
		5. Grades are a good estimate of the quality of learning that has taken place.
		6. Given the word "meaningful" as indicating your opinion of grading, rate it according to the strength of this opinion.
		7. The finer the grading system, the better it reflects the students' competence level.
	Motivating-Function	8. Grades provide a necessary incentive to hard work.
GN	Non-Learning Function	1. Grading should be abolished at the University level.
	Pressure-Anxiety	2. Grades induce too much worry.
	Non-Measuring-Function	3. Grades are no indication of what the student has learned in a course.
	Hate	4. Whoever put more grades into the scale should be hanged.
		5. I have nothing for grades but pure hate.
		6. It is grossly unfair to award a graduate student a "D" or an equivalent grade.
		7. Given the word "evil" as reflecting your opinion of grading, rate it to show the strength of this opinion.
	Dysfunction	8. Grading encourages students to cheat in examinations.

A few comments are necessary. In administering the battery the items would be thrown into some random order. Future revision will aim at ten items for each scale, at least three and at most four factors under each scale, and two or four items within each factor. The increase in the total number of items will hopefully lead to increase in validity, while the use of even number of items under each factor will make it convenient to compute split-half reliability coefficients.

TEST STATISTICS

In the present case where there were five alternative weighted responses the product moment correlation of item scores with the total scores in their appropriate scales may be used in determining items which belong to the Scale. But such coefficients are inflated since the item scores are also included in the scale scores. Even so these coefficients are displayed in Sub-appendix (h) together with the standard deviations for each item-variable, and also the inter-item correlations. The latter may safely be interpreted as <u>indices of belonging</u>. To facilitate their comprehension Table 8 summarizes the relevant data. It is worth noting that all coefficients are positive. Furthermore GP is the most homogeneous as its inter-item coefficients are all above .20. By the same standard GN is the poorest scale, and needs much revision.

TABLE 8

GROUPED FREQUENCIES, RANGE AND MEDIAN OF INTER-ITEM CORRELATIONS

Categories	EP (f)	EN (f)	GP (f)	GN (f)
.6000-.6999	1	1	1	-
.5000-.5999	4	6	7	1
.4000-.4999	16	12	29	12
.3000-.3999	24	22	17	15
.2000-.2999	15	20	12	20
.1000-.1999	6	5	-	13
Below .1000	-	-	-	5
Total (f)	66	66	66	66
Range	.159-.621	.129-.609	.228-.641	.004-.545
Median	.354	.320	.410	.277

Intercorrelation Among the Scales

Logically the total scores for the "positive" and "negative" scales should reveal an inverse relationship between them. But this may not be perfect since the "dimensions" are not necessarily on the same linear continuum. In fact the inverse relationship may be conceived to be an intrinsic property of the "negative" and "positive" dimension vectors. The absolute sizes of the coefficients as presented below also show an interesting pattern: the positive scales (EP-GP) and the negative scales (EN-GN) correlate more highly within their like-pairs than they do within unlike pairs

(EP-EN or EP-GN; similarly GP-GN or GP-EN). This may be interpreted as another evidence against bipolarity of the attitude factors. The correlation between "positive" and "negative" scales is negative; if the scales were on the same linear continuum, if they represented opposite ends of a bipolar factor then the absolute value of the correlation coefficient would be as close to unity as possible. The evidence of this study does not seem to support such a position. In the sample the correlations were as follows:

	EP	EN	GP	GN
EP	1.00			
EN	- .589	1.00		
GP	.796	- .562	1.00	
GN	- .550	.800	- .624	1.00

The directions of these coefficients agree with those illustrated on page 94.

Reliability of the Scales

An estimate of the reliabilities of the scales was computed by the Kuder-Richardson method. In the present case where responses are weighted the appropriate formula according to Magnusson (1966) is

$$r_{tt} = \frac{n}{n-1} \left(\frac{S_t^2 - \Sigma S_i^2}{S_t^2} \right)$$

where r_{tt} is the reliability coefficient (K-R_{20})

n is the number of observations

S_t^2 is the variance of the test

ΣS_i^2 is the sum of the item variances

The reliabilities shown below were based on this formula. The relevant data for calculations will be found in Sub-appendix (h).

EP	EN	GP	GN
.798	.791	.812	.746

INTERPRETATION OF THE SCORES

Ostensibly four scales make up this battery. However factor analysis has brought out sub-scales which are fairly easy to interpret. From the general instructions to the Questionnaire a value of 3 is to be assigned to a statement if "opinion hovers between agreement and disagreement equally." It will therefore follow that a mean score less than 3 or a mean score higher than 3 will be interpreted to indicate that the group or the individual is "low" or "high" on the particular dimension of attitude. The mean total score for a group of items may also be interpreted accordingly. Thus if there are four items in the sub-scale a mean total score of 12 would form the dividing line between the "lows" and the "highs" on the dimension reflected by that sub-scale.

The scheme for interpretation outlined implies a built-in meaning for the scores, and not a meaning to be determined with reference to any group. It seems logical that the meaning of scores should be similar to the Likert values as here

defined. The only assumptions are that the subject understands the instructions, and that he responds to the items honestly. These may be somewhat limited by "response-set" tendencies. The extent of such tendencies were not determined, but the percentage of respondents choosing each option shown in Table 9 would lead one to say that the effect of such sets may not have been very serious. The choices are fairly spread out except that respondents tend to avoid the high extreme value.

The above observations will now be illustrated for the try-out sample.

There are three factors in the EP Scale. In the first--the learning-function factor--there are five items; the mean total on these for the 573 observations is 12.8081. This places the group on the "low" end of this sub-scale with respect to their perception of examinations as a learning device. The mean item response on this and the other factors may be set out as follows:

Factor	Mean Item Response[a]	Range of Inter-item Correlations[b]
Learning Function	2.562	
Examination-Satisfaction	1.887	.159-.621
Motivating Function	2.584	

[a]The means for all items are given in the sub-appendix.

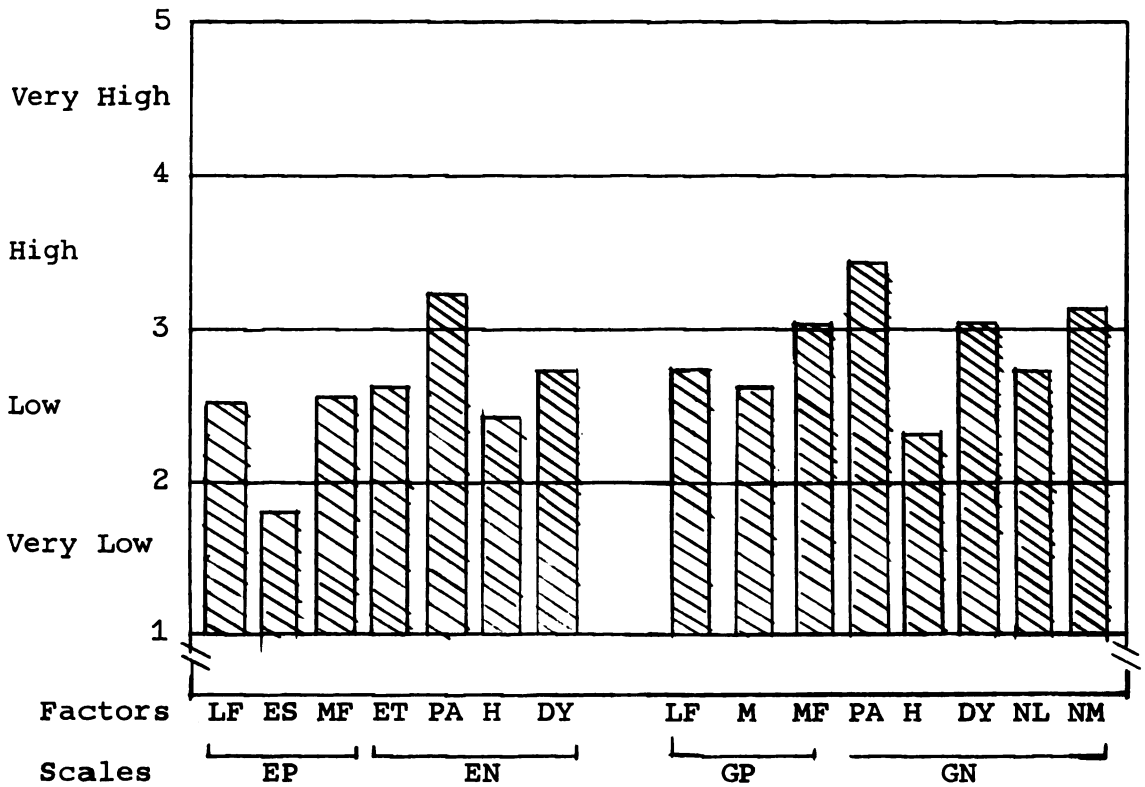
[b]These may be taken as estimates of the reliabilities of the factor scales.

These results also read "Low", or "Very Low", as on the Examination-Satisfaction factor.

The break-down of the other scales is as follows:

Scale	Factor	Mean Item Response	Range of Inter-item Correlations
EN	Examination-Type	2.613	.129-.609
	Pressure-Anxiety	3.328	
	Hate	2.531	
	Dysfunction	2.705	
GP	Learning-function	2.686	.228-.641
	Measuring-function	2.640	
	Motivating function	3.077	
GN	Pressure-Anxiety	3.415	.004-.545
	Hate	2.312	
	Dysfunction	3.059	
	Non-learning function (bipolar)	2.670	
	Non-measuring function	3.138	

The meaning that may be read into the above results is that the group tends to be "high" on the following factors: Pressure-Anxiety, Grade-Motivating function, Grade-Dysfunction and Grade-Non-Measuring function. On the other factors it is "low". The point needs emphasis. The scores for an individual (or group) on the Scales in this battery should be broken down into "factor" scores, and then interpreted in terms of "low" or "high" on the respective factors. The aim is to present a profile mapping of the individual in the defined attitude factors. Such a profile is presented in Figure 2 on the following page. The Pressure-Anxiety factors are prominent in both Examination and Grading scales, while the learning function factors are "low".



Key:
EP = Exam.-Positive
EN = Exam.-Negative
GP = Grade-Positive
GN = Grade-Negative
LF = Learning Function
ES = Exam. Satisfaction
MF = Motivating Function
M = Measuring (function)
ET = Exam. Type
PA = Pressure-Anxiety
H = Hate
DY = Dysfunction
NL = Non-learning Function
NM = Non-measuring Function

Figure 2. Attitude profile of the try-out sample (N = 573) Students attitudes towards Examination and Grading Scale Battery (SATEG SB).

A frequency count was made of respondents choosing each option, and converted into percentages. Table 9 shows these mean percentages under each factor sub-scale.

TABLE 9

MEAN PERCENT OF RESPONDENTS CHOOSING OPTION IN THE FACTOR SUB-SCALES

Scale and Factor	Likert-Pont Values 1	2	3	4	5
EP					
Learning Function	19	32	26	18	5
Exam. Satisfaction	45	27	14	7	2
Motivating Function	14	35	26	26	3
EN					
Exam. Type	14	39	22	19	5
Pressure-Anxiety	6	20	26	29	18
Hate	24	34	21	12	4
Dysfunction	10	36	30	18	5
GP					
Learning Function	21	24	25	21	7
Measuring Function	15	32	27	21	3
Motivating Function	10	21	28	34	7
GN					
Non-Learning Function	20	25	29	17	9
Pressure-Anxiety	5	19	22	34	19
Non-Measuring	7	28	24	25	16
Hate	29	27	22	11	6
Dysfunction	12	25	23	26	14

The picture shown may be easily comprehended if the options 4 and 5 are combined and summarily described as "high". (Similarly 1 and 2 may be combined and described as low.) On this basis the following statements may be made of this sample:

1) 22 percent are high on the learning function factor in the EP scale
2) 29 percent are high on the motivating function factor

In contrast only 9 percent are highly satisfied with examinations. The EN scale throws some light on this contrast. Here 47 percent are high on the Pressure-Anxiety factor and 23 percent on the Dysfunction factor.

A similar analysis may be made of the Grade Scales. For the GP scale the percentages on the high group are: learning function, 28; measuring function, 24; and motivating function 41. Thus the group perceives grades more as a motivating than as a measuring or learning device. The figures for the pressure-anxiety and dysfunction factors are 53 and 40 respectively. This would mean that more than half the sample perceive grades as generating pressure and anxiety, and about a half also feel grades perform no useful function.

IMPLICATIONS AND CONCLUSION

The manner of describing attitude on a psychological object as being "high" or "low" along specified attribute "factors" and the dimensions they support has some diagnostic value. At least it is a step beyond a global conception of attitude. Moreover it makes it comparatively easy to "control" attitude. Suppose for example that this scale is valid and that with its aid the attitude of a group in the learning-

function aspect of examinations is diagnosed to be "low"; an area is thus clearly specified for "treatment" should one desire to influence attitude on this positive dimension. In other words, control of attitude towards a psychological object becomes feasible if the anchors of the attitude are identified. It is reasonable to think that attitude change may be effected through some manipulation of the attributes of the psychological objects.

A cursory look at the pattern of the figures in Table 9 may lead one to suppose an inverse relationship between the learning and motivating function factors on the one hand and pressure-anxiety and dysfunction factors on the other. This suggests that the attitudes may be "changed" to be more "positive" if effort is concentrated on developing the learning and motivating function attributes of both examinations and grading. A general hypothesis may therefore be set out as follows: the more students perceive examinations and grading as promoting learning the less they will feel the pressure and anxiety which these twin aspect of the curriculum also generate, and hence the more positive will their attitudes be towards these objects, and consequently the higher the amount of learning that will take place. This general hypothesis may be broken up and tested, among others, in a program of construct validation of this scale battery.

The results of this study provide evidence which tend to agree with the theoretical model. Figure 3 reproduces the model with specific reference to the present study.

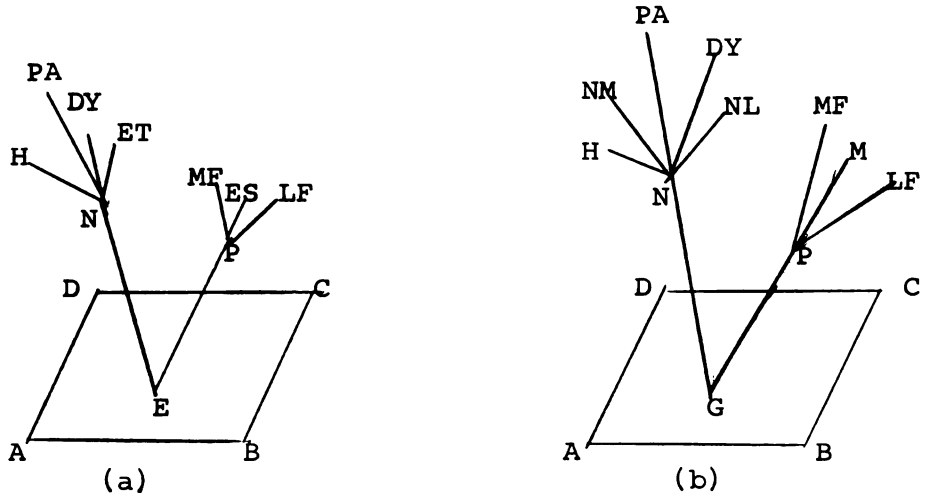


Figure 3. The attitude model with specific reference to Examinations (a) and Grading (b). (The reader is now familiar with the abbreviations used; the words they stand for are displayed in the Key to Figure 2; the general model is presented in Figure 1.)

In the figure ABCD still represents the attitude predispositional base, which remains the same for all attitudes of an individual. In fact both parts (a) and (b) would be shown on the same diagram; they are separated here to aid clear presentation. It should be noted that the growth points are now defined with reference to the attitude objects (E: Examinations; G: Grading). Furthermore the positive dimemsions (EP and GP) are parallel; so also the negative dimensions (EN and GN). The reader is reminded of the high and positive correlation between the scales in brackets, and of their loadings on the various factors discussed earlier. The last observation would lead one to suggest that positive attitudes, irrespective of the attitude objects

would correlate highly and positively with one another; similarly negative attitudes would correlate highly and positively.

The attribute vectors shown in Figure 3 represent the factors revealed in the factor analysis. The figure shows that the upward growth of attitude along each dimension-positive or negative is supported by the number and strength (reflected by length) of the attributes. The model and the evidence provided by this study would lead one to doubt that attitude is bipolar. A linear continuum model for attitude may not be appropriate.

An instrument like this can serve two purposes. It may be used for an attitude survey and in studies of relations between attitude and other variables. Furthermore it may be used to plan "treatment" measures to bring about attitude change. The traditional attitude measures do not seem to suggest this diagnostic and treatment use. In the writer's mind if social scientists survey attitude and always report it in the global form they are unwittingly perpetuating the attitude; and this may not always be desirable. If on the other hand their reports make evident the anchoring factors, someone's attention will be easily arrested to examine the basis of the attitude.

It must be added however that the model needs further supporting evidence to be worth considering. It is therefore suggested that the battery be used as a research

instrument--to investigate how stable the factors are across different student populations. Other workers may of course wish to test the model and the approach using different attitude objects.

SUB-APPENDIX (a)

THE OPEN-ENDED QUESTIONNAIRE

Course No. and Title:_____
Course Instructor:_____
Student's Name & No. (Optional)
_____
Date:_____

STUDENTS' OPINIONS AND ATTITUDES ON THE EXAMINATION-GRADING CONTROVERSY

Introduction and Instructions

The debate--"to examine or not to examine, to grade or not to grade"--is a very crucial one in college and university education today. To be democratic and also to help create a healthy atmosphere for carrying out our educational objectives it would be desirable for students to take part not only on this debate but in the formulation of policies on this issue. A survey is therefore being conducted to tap students' opinions and attitudes. Your response to the following questions will be of great importance in future decisions on examination and grading practices in this University. Consider it therefore a grand opportunity now offered you to influence policies in these areas. It is up to you in particular to utilize such a rare opportunity to express your views for your good and for the good of future generations of students.

To underline the importance of this survey to you in particular, you are to take this questionnaire home; respond to it independently and candidly and then return it to your instructor the following day.

Feel free to use the blank pages of this questionnaire to write as much as you like on any of the questions.

Thank you for your cooperation.

Do not write on this margin

Do not write on this margin

The Questions

1. How important are examinations in the instructional process? Defend your opinion.

2. How important is "grading" (involving the use of A, B-- or 0,1) in the instructional process? Defend your opinion.

3. Some say examinations and grading are a necessary evil while others believe they are an important aspect of the instructional process. How do you feel about these aspects of the curriculum? Defend your answer.

Do not write on this margin

4. What reactions have <u>you</u> had to the examinations you have taken in your college and university experience?

Do not write on this margin

5. In your college and university experience, what reactions have you had over your grades in particular and over the grading system in general?

6. Do you have suggestions for change that should be made in the examination practice at the college level? Defend your suggestions.

Do not write on this margin

7. Do you have suggestions for change that should be made in the grading practice at the college level? Defend your suggestions.

Do not write on this margin

8. Which examination type do you prefer more--the essay or the objective? State reasons for your preference.

9. Which of the following item types do you most prefer--True-False, Multiple Choice or Completion Type? State reasons for your preference.

Do not write on this margin

10. Which of the following item types do you <u>least</u> prefer--True-False, Multiple Choice or Completion Type? State reasons for your preference.

Do not write on this margin

11. Would you favor a <u>more</u> or a <u>less</u> emphasis on examinations at the university level? State your reasons for your answer.

12. Would you favor a <u>more</u> or a <u>less</u> emphasis on grading at the University level? State your reasons for your answer.

Do not write on this margin

13. It has been suggested that students should be involved directly and actively in the decisions determining their grades. Would you support this suggestion? State reasons for your opinion.

Do not write on this margin

14. If you can, suggest and defend concrete ways in which students might be directly and actively involved in the determination of their grades.

15. Would you, or would you not, support a student motion urging the complete abolition of examinations and grading at the college level? State reasons for the position you take.

SUB-APPENDIX (b)

SCHEME FOR THE CONTENT ANALYSIS

(i) Coding and Categorization

Description of Item	Coding Symbol
Positive[1] direction of attitude toward examination	+ ex
Negative[2] direction of attitude against examination	- ex
Positive direction of attitude toward grading	+ gr
Negative direction of attitude against grading	- gr
Content Categories:	
1. Statement of function (e.g., feedback; stifles learning)	A
2. Statement of preferences--either in direct answer to "which . . . prefer?" or implied in statement	B
3. Statement expressing or implying emotion (e.g., very important, less emphasis)	C
4. Statement offering suggestions directly (e.g., term paper)	D
Question number--Use Roman numerals	I, II...XV
Respondent: assigned Arabic numerals to be written after the course number, and separated by a colon:--	ED200:4

(ii) General Procedural Steps in the Analysis

1. Read through the response to each question.
2. Re-read, and underline significant words, etc., which may be put into one of the content categories, and append the appropriate code symbol.
3. Judge direction of attitude as either positive or negative and append the code (+ ex, for example) beside the content code of every underlined word, etc.
4. Transfer the coding symbols to the right margin (use the left margin for writing comments, if any).
5. On the outline summary blank provided prepare the "Summary of Analysis" table (as shown below) and transfer the results of the analysis.
6. All work is to be done on pencil.

---

[1]Examples of words, etc.: "feedback"; "very important"

[2]Examples: "stifles learning"; "less emphasis"

(iii) Specific Hints on the Analysis of Each Question

Item No.*	Hints
1.	Perhaps this Q. will prove the best stimulus eliciting responses illustrating "statement of function"--e.g., 1. motivates learning 2. assesses performance 3. reveals weaknesses in learning 4. guides learning.
2.	Perhaps best stimulus eliciting 1) incentive to study, work hard; 2) reward. Category A or C may abound, but others not excluded. This comment also applies for number 1 and other items.
3.	On the surface this Q seems a repetition of number 1 and 2 but a new stimulus is subtly introduced in "necessary evil." If respondents agree with this stimulus then the direction of their attitude tends to be negative. Look out for attitudinal and emotional overtones.
4.	Some reactions will reflect positive attitudes, others negative. Rate (judge) each key word, etc., appropriately. Perhaps "statement on efficiency/ inefficiency" will be elicited--(Category C).
5.	Same remarks as in number 4.
6.	The attitude object is written examination.* The following therefore reflect negative attitude (-)Ex 1. Oral exams 2. Term papers 3. Reports of projects, etc. On the other hand the following are positive 1. More emphasis on essay exams 2. More emphasis on objective exams 3. More quizzes, etc. (open-book, take-home) *Involving a series of test items--objective or essay, taken in class or at home, closed-look or open-book.
7.	The attitude object is the grading system involving at least three levels--whether letters or numerals, and "GPA." Therefore suggestion of 1. Pass--Fail 2. Pass--No credit, etc. show negative attitude. Positive attitude is reflected by 1. a finer system 2. a broader system 3. a narrower, etc.

*See Sub-appendix (a).

Item No.	Hints
8.	Main response here is in category 8 expressed "statement of preference"; direction is positive. Judge direction of stated reasons and their categories separately. Score negative the response 'none'.
9.	Same remarks as for number 8.
10.	Main response is in stated preference category (8) and direction is negative. Judge category and direction of reasons separately.
11.	Response of "more emphasis" reflects "clear indication of satisfaction--category C; direction is positive. On the contrary "less emphasis" is negative and in category C--Judge reasons separately.
12.	Same remarks as for number 11.
13.	Mere Yes or No response is not scored. Base direction and category on the reasons--some will be pro-grading, others against, e.g., mutual discussion of grades determination is +ve. Pass-Fail is -ve.
14.	Any suggestion reflects positive attitude and is scored under category D, e.g., discussion with students of what goes into the grade.
15.	"Complete abolition" has emotional-attitudinal overtone and so response is to be scored in category C. The nature of reasons helps to determine direction also. Support shows negative attitude.

(iv) Summary of Analysis Table

(An actual entry is provided from Course No. and Respondent's No. a respondent--Student ED450:1)

	+		-	
	Examination	Grading	Examination	Grading
A	[1] 1. Feedback to instructor		1. Little more than the instructor's scape goat	1. Stifles learning
B	2. Essay-type 3. Completion type		2. Multiple-Guess	
C	4. Essay is personalized 5. Essay allows one's expression 6. Completion is direct	1. Possibly it is important in the Graduate School	3. Full of errors 4. Inadequate coverage 5. Idiocy of choosing 6. Less emphasis 7. Entirely unnecessary 8. Abolish	2. No importance for undergraduates 3. Brings pressure 4. What does it show? 5. Fosters cramming 6. Abolish
D		2. Award grades on papers and Class Discussions		7. Pass-Fail
Column[2] Score	6	2	8	7
Placement[3]			x	x
Comments[4] if necessary				

[1]The cell entries are the significant words, etc., marked each entry is to be numbered and the item response number indicated. Number down a column only.

[2]To be determined by a count of all entries.

[3]This is determined by the direction of the difference between the column scores under "examination" and "grading" where the difference is zero place according to response to item XV.

[4]Note entries thought to be useful for an attitude statement.

SUB-APPENDIX (c)

INSTRUCTIONS FOR JUDGMENT AND Q-SORTING OF STATEMENTS

1. Name:_____
2. Academic Qualifications:_____
3. Present Degree Program:_____
4. Date_____

Examinations refer to classroom achievement testing involving the administration of quizzes, mid-terms and finals.

Grading refers to evaluation of students' academic performance in examinations and/or other aspects of the curriculum by using letters (e.g., A, B, . . . F) or numbers (e.g., 4, 5, 4.0 . . . 0.0; etc.) to classify students according to their relative achievement levels, provided that the system also involves the report of "grade-point-averages."

This exercise involves two stages:

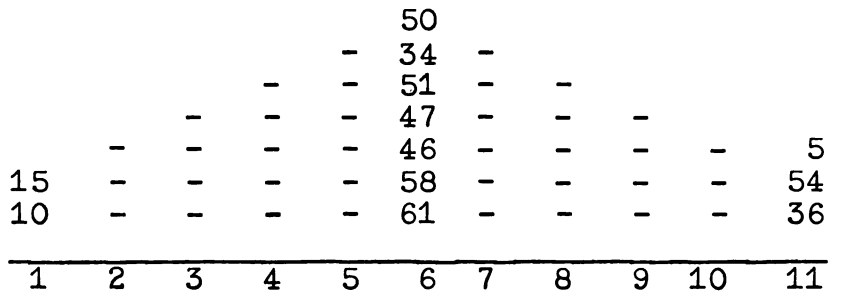
(a) Judgment--
in which you say whether the statement reflects a favorable positive attitude (P) or an unfavorable negative attitude (N) towards the named object.

and either

(b) Rating--
in which you assign the statement a place on an eleven-point scale in which 1 represents the lowest degree and 11 the highest degree of the judged direction.

or

(c) Sorting--
in which you (1) group the statements under two main headings: EXAMINATIONS, GRADING, (2) form sub-groups of positive and negative statements under each main group, (3) arrange the statements in each sub-group on an eleven-point scale in which 1 represents the lowest degree and 11 the highest degree of the judged direction. To do this, view all the statements in the sub-group as a whole; then decide which among them will have the lowest value and place it (or them) above the value 1; further decide which has the highest value and place it (or them) above the value 11. Finally, arrange the other statements and place them above any of the values 2 to 10 as you judge them appropriate. Your final results will look something like this:



It is advisable to use rough paper at first. Summarize your final results in the appropriate spaces provided below.

Summary of Results of Judgment and Sorting

1 2 3 4 5 6 7 8 9 10 11		1 2 3 4 5 6 7 8 9 10 11
Positive		Negative

EXAMINATIONS

1 2 3 4 5 6 7 8 9 10 11		1 2 3 4 5 6 7 8 9 10 11
Positive		Negative

GRADING

Pile up vertically above each scale value the number of statements assigned to the scale value.

SUB-APPENDIX (d)

"JUDGMENT" ON THE MEDIAN VALUES OF THE ORIGINAL 65 ITEMS

Statement	Judgment	Value
1. Without examinations most students will not study.	EP	3
2. Examinations are the best means for motivating college students to learn.	EP	8
3. The taking of examinations brings about a highly valued learning experience.	EP	9
4. Of all teaching devices, examinations provide the most useful feedback	EP	6
5. Examinations enforce my desire to learn	EP	6
6. Given the word "favorable" as referring to your feeling about examinations, rate it to indicate the degree of this feeling.	EP	6
7. Grades stifle learning.	GN	8
8. Grading encourages students to cheat in examinations.	GN	7
9. Grades sometimes make me feel helpless and insecure.	GN	5
10. There should be less emphasis on grading at the university level.	GN	4
11. Examinations force students to cram facts without real understanding.	EN	4
12. There is very little of instructional value in examinations.	EN	5
13. Examinations are the scapegoat for most instructors.	EN	3
14. Examinations are redundant in the educational process at the university level.	EN	5
15. Grades are very effective for indicating students' achievement of the course objectives.	GP	8
16. Grades surpass in usefulness other measures of academic progress.	GP	8
17. Grades are a good estimate of the quality of learning that has taken place.	GP	6
18. Grades provide a necessary incentive to work hard.	GP	7
19. Grades differentiate the serious-minded from the care-free student.	GP	5
20. Given the word "meaningful" as indicating your opinion of grading, rate it according to the strength of this opinion.	GP	6

Statement	Judgment	Value
21. Objective examinations are nothing more than a guessing game.	EN	5
22. Examinations provide the student a frustrating experience.	EN	5
23. I resent the pressure which examinations bring on me.	EN	6
24. Examinations should be abolished at the university level.	EN	9
25. Most examinations pose stupid and ridiculous questions.	EN	8
26. Given the word "useless" as indicating your opinion of examinations rate it according to the strength of such opinion.	EN	7
27. The grading system should be an integral part of the curriculum in higher education.	GP	9
28. For the student, grades are a desirable aid to self-evaluation.	GP	7
29. Abolition of grading would jeopardize learning at the university level.	GP	8
30. Grading is a necessity if standards have to be maintained in university education.	GP	6
31. Examinations provide the most satisfactory means for assessing learning.	EP	7
32. Examinations should be an indispensable feature of the university curriculum.	EP	6
33. I am satisfied with the university examinations system.	EP	3
34. Without examinations, academic standards would fall.	EP	6
35. Grades induce too much worry.	GN	6
36. Grading should be abolished at the university level.	GN	10
37. Most students' interests are diverted from learning to grades--as a goal.	GN	8
38. Grades prove nothing.	GN	5
39. Grades are no indication of what the student has learned in a course.	GN	6
40. Given the word "evil" as reflecting your opinion of grading, rate it to show the strength of this opinion.	GN	6
41. Examinations should be given more emphasis in the university curriculum.	EP	6
42. I enjoy taking examinations.	EP	5
43. The discipline of examinations is vital to learning.	EP	8
44. Examinations make study exciting.	EP	7
45. Abolition of examinations will in the long run lead to chaos in graduate education.	EP	7

Statement	Judgment	Value
46. Given the word "acceptable" as reflecting your feelings on examination, rate it to indicate the intensity of your feelings.	EP	4
47. I would campaign vigorously against any attempt to abolish grading at the university level.	GP	11
48. Without grading the motivational function of examinations would be impaired.	GP	5
49. In general I have no complaint against my grades.	GP	1
50. Guaranteeing graduate students an "A" or a "B" in a course is insulting to them.	GP	2
51. The finer the grading system the better it reflects the students' competence level.	GP	6
52. There is no conflict between working for grades and gaining knowledge.	GP	2
53. Given the word "relevant" as describing your opinion of grades, rate it to show the strength of this opinion.	GP	5
54. In general, examinations appeal to rote memory.	EN	1
55. Very little learning, if any, is derived from taking examinations.	EN	8
56. Examinations generate too much anxiety.	EN	7
57. Examinations are nothing more than trickery.	EN	10
58. Grades are of no importance in the educational process at the University level.	GN	7
59. The examination system is entirely lacking in precision.	EN	5
60. Given the phrase "a farce" as indicating your opinion of examinations, rate it to show the strength of this opinion.	EN	8
61. In my experience as a university student, examinations are the instructors makeshift, without any real value.	GN	5
62. I have nothing for grades but pure hate.	GN	11
63. Whoever put more grades into the scale should be hanged.	GN	10
64. It is grossly unfair to award a graduate student a "D. or an equivalent grade.	GN	7
65. Given the word "inadequate" as reflecting your opinion of grading, rate it to show the strength of this opinion.	GN	6

SUB-APPENDIX (e)

COMPOSITION OF THE MAIN TRY-OUT SAMPLE

College[1]	Department	Undergraduate Level Course No.	Undergraduate Level Returns	Graduate Level Course No.	Graduate Level Returns	Totals
Arts and Letters	Art	355	26	802	15	41
	English	402	35	811	13	48
Business	Economics	324	25	811	31	56
Communication Arts	Journalism	310	19	800	11	30
Education	Education	325c	35	867	97	132
Natural Science	Botany	301	55	943	6	61
	Chemistry	351	22	811	26	48
	Mathematics	321	34	847	17	51
	Physics	395	32	837	39	71
Social Science	Psychology	310	22	800	15	37
Totals			305		280	585

[1]Other Colleges of the University not directly sampled include (1) Agriculture, (2) Engineering, (3) Home Economics, (4) Human Medicine and Veterinary Medicine. The loss is apparent, not real. Some students in the department of Economics major in Agriculture, some of those in Mathematics and Physics major in Engineering. In Education, Home Economics majors are quite common, and similarly majors in Medicine are to be found in the Botany, Chemistry and Physics departments. In short, the sample is fairly representative of the population of students in Michigan State University.

SUB-APPENDIX (f)

THE FACTOR PATTERNS AND LOADINGS

Preliminary Notes

1. Arbitrary conditions for a factor to be considered "significant": either

   (a) five percent (or more) of the common variance is accounted for by the factor, or
   (b) at least three variables are "significantly" loaded on the factor.

   (Here a significant loading $\geq$/.40/)

   Non-significant loadings are not recorded.

2. Excepting those in parentheses ( ) recorded loadings are the highest in the row in relation to other loadings of the variable on subsequent factors.

3. Note that the decimal point precedes every "loading" entry.

4. Loadings which are not the highest in the row are enclosed in brackets.

Variable No.	Description	Quartimax 1					Quartimax 2				
		1	2	3	4	5	1	2	3	4	5
1	EP	-923					-928				
2	EN	701			474		(518)	752			
3	GP	-943					-944				
4	GN	712	526	-408			(538)	773			
5	EP_2	-666					-618				
6	EP_3	-565					-562				
7	EP_4	-547					-609				
8	EP_5	-599					-568				
9	GN_6	591					452				
10	GN_7			-422							
11	EP_8								535		
12	EN_9				473					467	
13	EN_{10}	529								570	
14	GP_{11}	-585					-534				
15	GP_{12}	-570					-499				
16	GP_{13}	-642					-706				
17	GP_{14}	-655					-614				
18	EN_{15}				617			410			
19	EN_{16}			-486				549			
20	EN_{17}			-622				512			
21	EN_{18}	684					615				
22	EN_{19}	616					553				
23	GP_{20}	-701					-721				
24	GP_{21}	-743					-657				
25	GP_{22}	-728					-703				
26	GP_{23}	-761					-732				
27	EP_{24}	-669					-686				
28	EP_{25}	-692					-794				
29	EP_{26}	-632					-784				
30	GN_{27}	(408)		-664				467			
31	GN_{28}	737					617				
32	GN_{29}	(451)		-456			415				
33	GN_{30}	528									
34	GN_{31}	486					410	570			
35	EP_{32}	-541									
36	EP_{33}	-701					-709				
37	EP_{34}	-471					-546				
38	EP_{35}	-628					-588				
39	GP_{36}	-657					-509				
40	GP_{37}	-552					-580				
41	GP_{38}					-501	-484				
42	GP_{39}	-722					-672				
43	EN_{40}			-727				534	-561		
44	EN_{41}				564			590			
45	GN_{42}	593						509			
46	EN_{43}							480			
47	EN_{44}	536	421					634			
48	EN_{45}	554						632			
49	GN_{46}	(458)	489					683			
50	GN_{47}		558					572			
51	GN_{48}		483					400			
52	GN_{49}						466				
Percent of Variance		33.05	5.84	6.15	4.45	3.19	27.88	13.17	3.7	3.36	
		4 factors significant					2 factors significant				

Varimax 1

		1	2	3	4	5	6	7	8
1	EP*	-748			-519				
2	EN*			-456			-661		
3	GP*	-706				-432			-439
4	GN*		544						531
5	EP	-782							
6	EP	-434							
7	EP				-607				
8	EP				-519				
9	GN								542
10	GN							-556	
11	EP				-580				
12	EN						-664		
13	EN						-424		
14	GP								
15	GP					-532			
16	GP	-536							
17	GP								-465
18	EN						-680		
19	EN			-707					
20	EN			-716					
21	EN	474							
22	EN						-512		
23	GP	-497						-443	
24	GP	-584							
25	GP	-775							
26	GP	-758							
27	EP	-574							
28	EP	-693							
29	EP	-764							
30	GN			-597				453	
31	GN	523						465	
32	GN						-429	498	
33	GN							623	
34	GN								
35	EP	-490					415		
36	EP	-648							
37	EP								
38	EP	-613							
39	GP	-547							
40	GP	-544							
41	GP					-734			
42	GP	-447						-434	
43	EN			-756					
44	EN						-644		
45	GN							456	
46	EN						-429		
47	EN		(487)				-490		
48	EN		(427)				-484		
49	GN		637						
50	GN		727						
51	GN		612						
52	GN		470					-544	
Percent variance		17.08	6.5	6.39	5.58	4.22	8.37	3.86	7.93

8 factors are significant

*Scale

Varimax 2

		1	2	3	4	5	6	7
1	EP*	-745					-551	
2	EN*		443		698			
3	GP*	-838						
4	GN*		586		426			
5	EP						-457	
6	EP						-519	
7	EP						-534	
8	EP						-428	-410
9	GN							499
10	GN							586
11	EP			448			-547	
12	EN				564			
13	EN				654			
14	GP					-610		
15	GP					-601		
16	GP	-610						
17	GP					-553		
18	EN				511			
19	EN							
20	EN			-454				
21	EN				493			
22	EN				454			
23	GP	-660						
24	GP	-529						
25	GP	-727						
26	GP	-760						
27	EP	-523						
28	EP	-692						
29	EP	-737						
30	GN			-713				
31	GN	577						
32	GN				423			
33	GN					488		
34	GN		421		428			
35	EP							
36	EP	-654						
37	EP	-483						
38	EP	-592						
39	GP	-544						
40	GP	-611						
41	GP	-512						
42	GP	-545						
43	EN			-721				
44	EN				587			
45	EN		513					
46	EN					488		
47	EN		474		463			
48	EN		548		418			
49	GN		627					
50	GN		735					
51	GN		569					
52	GN							
Percent variance		18.49	7.72	5.35	9.35	5.97	5.27	3.84

7 factors are significant

*Scale

Variable No.	Description	Quartimax 3									
		1	2	3	4	5	6	7	8	9	10
1	EP	912									
2	EN	613	498	-409					-400		
3	GP	-943									
4	GN	644	589								
5	EP_2	-644									
6	EP_3	-542									
7	EP_4	-569					404				
8	EP_5	-576									
9	GN_6	542									
10	GN_7									598	
11	EP_8						667				
12	EN_9								-703		
13	EN_{10}	(420)							-555		
14	GP_{11}	-536			458						
15	GP_{12}	-510			461						
16	GP_{13}	-685									
17	GP_{14}	-621			412						
18	EN_{15}								-448		
19	EN_{16}			-588							
20	EN_{17}			-631							
21	EN_{18}	669									
22	EN_{19}	575									
23	GP_{20}	-724									
24	GP_{21}	-698									
25	GP_{22}	-748									
26	GP_{23}	-792									
27	EP_{24}	-666									
28	EP_{25}	-744									
29	EP_{26}	-721									
30	GN_{27}			-690							
31	GN_{28}	707									
32	GN_{29}	444									
33	GN_{30}	(460)			-480						
34	GN_{31}	(465)	470								
35	EP_{32}	-424						-511			
36	EP_{33}	-703									
37	EP_{34}	-484					443				
38	EP_{35}	-613									
39	GP_{36}	-588									
40	GP_{37}	-585									
41	GP_{38}	(401)				610					
42	GP_{39}	-684									
43	EN_{40}			-758							
44	EN_{41}		463								
45	GN_{42}	528	(401)								
46	EN_{43}		491								
47	EN_{44}	(485)	592								
48	EN_{45}	(450)	569								
49	GN_{46}		649								
50	GN_{47}		677								
51	GN_{48}		532								
52	GN_{49}	(401)	406								
Percent variance		30.35	7.81	5.24	2.92	2.43	2.77	2.48	3.4	2.38	2.40

6 factors are "significant"

Variable No.	Description	Varimax 3									
		1	2	3	4	5	6	7	8	9	10
1	EP*	-726									
2	EN*		542	-457							
3	GP*	-707			615						
4	GN*			(-441)		-570					
5	EP	(-455)							-577		
6	EP	-400									
7	EP								-596		
8	EP								-533		
9	GN										
10	GN									-453	
11	EP						578				
12	EN									-633	
13	EN									-680	
14	GP				659						
15	GP				626						
16	GP	-526							(-398)		
17	GP				581						
18	EN		702								
19	EN			-579							
20	EN			-711							
21	EN	413									
22	EN										
23	GP	-521			429						
24	GP	-503			417						
25	GP	-757									
26	GP	-772									
27	EP	-565									
28	EP	-719									
29	EP	-791									
30	GN			-745							
31	GN	515									
32	GN										
33	GN				-446						
34	GN					-453					
35	EP						424				
36	EP	-662									
37	EP						631				
38	EP	-634									
39	GP	-591									
40	GP	-533									
41	GP				512						
42	GP	(-430)			549						
43	EN			-780							
44	EN		664								
45	GN										
46	EN							520			
47	EN					-439		434			
48	EN					-433					
49	GN					-643					
50	GN					-737					
51	GN					-596		535			
52	GN										
Percent variance		16.08	5.43	7.06	7.69	6.54	3.42	3.95	4.91	4.52	2.58

9 factors are "significant"

SUB-APPENDIX (g)

A COMPARISON OF THE VARIMAX FACTORS ACROSS THE THREE SAMPLES AND THE FOUR SCALES

Preliminary Notes

1) The cell entries are the variable numbers as already given in Sub-appendix f.

2) The variables recorded have significant loadings as defined in Sub-appendix f.

3) Variables common in a factor across the three samples are underlined.

Scale	Comment
EP	1. It would appear factor 4 of sample 1 and factor 6 of sample 2 and factor 8 of sample 3 are the same. 2. In all samples EP has been broken into two factors.
EN	Factor 6 (sample 1) is probably the same as factor 4 (sample 2). It is this factor that is separated into four in sample 3.
GP	Samples 1 and 2 bring out quite similar factors. But sample 3 groups into one factor (number 4) what sample 1 has as two factors (numbers 5 and 8).
GN	Factor 2 in both samples 1 and 2 is identical with factor 5 in sample 3.

Conclusion

The factor patterns across samples are similar though not in perfect agreement. The first factors are identical. Some of the factors merely switch positions. The evidence shows they are fairly stable and reliable. Attempts at naming these factors will be found in the relevant section of the main text.

		Factors								
Scale	Sample	1	2	3	4	5	6	7	8	9
	Sample 1	1*,5,6,27, 28,29,35, 36,38	none	none	1*,7,8, 11	none	none	35	none	none
EP	Sample 2	1*,27,28, 29,36,37, 38	none	11	none	none	1*,5,6 7,8	8	none	none
	Sample 3	1*,6,27, 28,29,36, 38	none	none	none	none	11,35,37	none	5,7,8	none
	Sample 1	21	none	2*,19, 20,43	none	none	2*,12,13, 18,22,44, 46,47,48	none	none	none
EN	Sample 2	none	2*,47, 48	20,43	2*,12,13, 18,21,22, 44,47,48	46	none	none	none	none
	Sample 3	21	2*,18, 44	2*,19, 20,43	none	47,48	none	46,47	none	12,13
	Sample 1	3*,16,23, 24,25,26, 39,40,41, 42	none	none	none	3*,15, 41	none	none	3*,17, 23,42	none
GP	Sample 2	3*,16,23, 24,25,26, 39,40,41, 42	none	none	none	14,15, 17	none	none	none	none
	Sample 3	3*,16,23, 24,25,26, 39,40	none	none	3*,14,15, 17,23,24, 41,42	none	none	none	none	none
	Sample 1	31	4*,49, 50,51 52	30	none	none	none	10,32 52	4*,9, 30,31, 32,33, 45	
GN	Sample 2	31	4*,34, 45,49, 50,51	30	4*,32,34	33	none	9,10	none	none
	Sample 3	31	none	30	33	4*,34, 49,50 51	none	51	none	10
Per.	Sample 1	17.08	6.5	6.39	5.58	4.22	8.37	3.86	7.93	
Var.	Sample 2	18.49	7.72	5.35	9.35	5.97	5.27	3.84	----	
	Sample 3	16.08	5.43	7.06	7.69	6.54	3.42	3.95	4.91	4.52

*Variables with asterisks are the four scales.

SUB-APPENDIX (h)

SCALE-ITEM AND INTER-ITEM CORRELATIONS

(a) Scale-item and Inter-item Correlations

The EP Scale

Variable Number	*$S_i^{(a)}$	r_{it} Scale EP	$r_{ii'}$ Variable Numbers 5	6	7	8	11	27	28	29	35	36	37	38
5	1.081	.660												
6	1.114	.629	.445											
7	1.229	.657	.468	.329										
8	1.062	.630	.440	.398	.458									
11	1.014	.437	.176	.220	.278	.291								
27	1.047	.690	.435	.418	.355	.371	.120							
28	1.152	.758	.433	.418	.405	.366	.232	.587						
29	1.188	.710	.424	.365	.365	.308	.198	.493	.620					
35	0.816	.485	.228	.218	.235	.187	.158	.310	.330	.280				
36	1.109	.723	.426	.374	.388	.365	.194	.453	.588	.541	.342			
37	1.004	.587	.269	.253	.393	.329	.320	.325	.354	.301	.315	.389		
38	1.121	.625	.311	.282	.267	.291	.208	.406	.483	.505	.310	.428	.276	

The EN Scale

Variable Number	*$S_i^{(a)}$	r_{it} Scale EN	$r_{ii'}$ Variable Numbers 12	13	18	19	20	21	22	43	44	46	47	48
12	1.074	.496												
13	1.035	.559	.384											
18	1.116	.568	.246	.285										
19	1.075	.577	.182	.208	.296									
20	1.233	.637	.185	.262	.237	.461								
21	1.237	.712	.318	.426	.357	.273	.440							
22	1.126	.682	.319	.393	.324	.283	.345	.545						
43	1.173	.596	.128	.207	.257	.430	.539	.331	.202					
44	1.003	.689	.287	.287	.491	.415	.287	.416	.420	.337				
46	1.144	.575	.185	.183	.217	.258	.272	.282	.324	.312	.361			
47	1.207	.737	.272	.345	.338	.301	.363	.480	.535	.339	.500	.523		
48	1.059	.711	.302	.307	.313	.294	.349	.455	.469	.317	.501	.471	.608	

*(a) Standard deviation.

The GP Scale

Variable Number	*(a) s_i	r_{it} Scale 3	$r_{ii'}$ Variable Numbers										
		GP	14	15	16	17	23	24	25	26	49	40	41	42
14	1.055	.628												
15	1.051	.598	.477											
16	1.115	.680	.357	.331										
17	1.005	.679	.454	.456	.437									
23	1.117	.751	.414	.371	.480	.491								
24	1.102	.729	.458	.410	.440	.433	.533							
25	1.200	.714	.349	.291	.499	.404	.480	.481						
26	1.192	.735	.332	.344	.518	.394	.515	.516	.640					
39	1.173	.602	.325	.248	.301	.312	.428	.406	.427	.405				
40	1.227	.626	.286	.240	.377	.335	.432	.409	.420	.426	.323			
41	1.228	.515	.239	.293	.283	.241	.336	.286	.237	.276	.227	.337		
42	1.062	.736	.431	.392	.406	.583	.545	.526	.434	.466	.406	.408	.342	

The GN Scale

Variable Number	*(a) s_i	Scale 4												
		GN	9	10	30	31	32	33	34	45	49	50	51	52
9	1.129	.606												
10	1.249	.418	.264											
30	1.178	.590	.289	.218										
31	1.229	.703	.457	.211	.405									
32	1.053	.558	.312	.246	.430	.422								
33	1.194	.609	.395	.124	.338	.415	.418							
34	1.195	.676	.393	.212	.312	.447	.280	.407						
45	1.129	.652	.360	.148	.276	.544	.285	.381	.431					
49	1.105	.662	.275	.141	.321	.389	.263	.330	.481	.418				
50	1.348	.494	.168	.121	.148	.210	.043	.133	.283	.277	.463			
51	1.380	.375	.074	.004	.131	.105	.040	.096	.143	.168	.241	.278		
52	1.329	.587	.271	.185	.269	.326	.311	.272	.332	.321	.312	.218	.195	

*(a)Standard deviation.

SUB-APPENDIX (i)

SUMMARY OF MEAN SCORES AND STANDARD DEVIATIONS

Variable	Mean	S.D.	Variable	Mean	S.D.
1	29.3141	8.3172	27	2.5864	1.0477
2	33.2496	8.5111	28	2.5812	1.1528
3	33.4293	9.0263	29	2.6300	1.1886
4	34.0995	8.3214	30	3.4154	1.1780
5	2.6405	1.0813	31	2.6702	1.2298
6	2.8098	1.1149	32	3.8534	1.0533
7	2.5323	1.2298	33	3.1379	1.1949
8	2.7469	1.0620	34	2.4293	1.1951
9	2.7243	1.1290	35	1.7260	0.8165
10	3.0593	1.2491	36	2.6667	1.1095
11	2.2129	1.0147	37	1.8866	1.0040
12	2.7016	1.0748	38	2.3438	1.1219
13	2.7086	1.0354	39	2.1798	1.1731
14	2.6405	1.0551	40	3.0855	1.2271
15	2.3665	1.0511	41	2.7853	1.2284
16	3.0768	1.1152	42	2.8604	1.0625
17	2.7784	1.0051	43	3.5794	1.1739
18	2.6126	1.1163	44	2.1850	1.0030
19	3.2216	1.0756	45	2.4154	1.1296
20	3.1815	1.2330	46	2.9564	1.1440
21	2.5759	1.2378	47	2.5462	1.2079
22	2.4904	1.1265	48	2.5166	1.0594
23	2.7958	1.1171	49	2.0838	1.1053
24	3.0175	1.1020	50	2.1518	1.3484
25	2.8569	1.2008	51	2.5812	1.3802
26	3.0227	1.1924	52	3.5864	1.3296

SUB-APPENDIX (j)

VARIMAX ROTATION ANALYSIS--MAIN TRY-OUT DATA

	1	2	3	4	5	6	7	8	9	10
1	-0.7262	-0.1023	0.1671	0.2302	0.0518	0.3721	-0.1970	-0.3933	0.1321	-0.0956
2	0.2739	0.5425	-0.4576	-0.1611	-0.3438	-0.1013	0.2556	0.2056	-0.3664	-0.0266
3	-0.7071	-0.1023	0.1118	0.6159	0.1597	0.0963	-0.0354	-0.1550	0.1262	0.0571
4	0.2992	0.1501	-0.4411	-0.3030	-0.5703	-0.0322	0.2779	0.1295	-0.2866	-0.2528
5	-0.4559	-0.0345	0.0284	0.1765	-0.0222	0.0247	-0.0845	-0.5770	0.1656	-0.0721
6	-0.4008	-0.0010	0.1290	0.1950	-0.0077	0.1047	-0.1626	-0.3342	0.1758	-0.3526
7	-0.3110	-0.1057	0.1205	0.1441	0.0274	0.3301	-0.0871	-0.5960	-0.0313	0.1554
8	-0.2501	-0.0613	0.2178	0.2980	0.1433	0.1882	0.0100	-0.5331	0.1874	-0.0425
9	0.2510	0.0248	-0.2428	-0.3672	-0.1720	0.1802	0.0570	0.2701	-0.3571	-0.1761
10	-0.0022	-0.2253	-0.3506	-0.1605	-0.0832	0.1260	0.2643	0.0285	-0.4539	0.1038
11	-0.1020	-0.0641	0.3031	0.0957	0.1043	0.5787	-.1506	-0.2369	-0.1052	-0.2803
12	0.0806	0.3776	0.0206	-0.0671	0.0159	-0.1714	0.0983	0.0312	-0.6338	0.0091
13	0.2254	0.1962	-0.1255	-0.0951	-0.1481	-0.0014	-0.0246	0.0753	-0.6806	-0.0320
14	-0.2003	0.0047	0.1130	0.6593	0.0974	0.1534	-0.1268	-0.1499	0.0810	0.0428
15	-0.2067	-0.1079	0.1400	0.6262	0.0539	0.0619	-0.3160	-0.0952	0.0262	-0.0978
16	-0.5269	-0.0114	-0.0526	0.3318	0.1193	-0.0729	-0.0565	-0.3984	-0.0032	0.1411
17	-0.3032	-0.2388	0.1031	0.5815	0.0088	0.0238	-0.1322	-0.2228	0.1034	0.1064
18	0.1167	0.7024	-0.1586	-0.1250	-0.0571	-0.0387	-0.0177	-0.0052	-0.1624	-0.0636
19	0.0618	0.3741	-0.5790	-0.0777	-0.1284	-0.0602	0.0663	0.0581	-0.0392	0.1918
20	0.1730	0.1206	-0.7110	-0.0316	-0.1727	-0.0268	0.0459	0.2204	-0.1410	0.1070
21	0.4138	0.2549	-0.2704	-0.1573	-0.2015	-0.1114	0.0040	0.2990	-0.3492	-0.2281
22	0.3308	0.3069	-0.0601	-0.1043	-0.3109	-0.1233	0.2028	0.3126	-0.3346	-0.0410
23	-0.5212	-0.0401	0.1521	0.4292	0.1884	0.0711	0.0090	-0.1306	0.1638	0.1614
24	-0.5033	-0.1136	0.1652	0.4175	0.1473	0.1091	-0.0405	-0.1464	0.1091	-0.0176
25	-0.7572	-0.0377	0.0807	0.1810	0.0875	-0.0746	-0.0235	-0.1322	-0.0272	0.1517
26	-0.7726	-0.0929	0.1143	0.1882	0.0917	-0.0853	0.0193	-0.1532	0.0187	0.0303
27	-0.5654	-0.1309	0.0976	0.2437	-0.0910	0.0407	-0.3590	-0.1671	0.1073	-0.1579
28	-0.7194	-0.1462	0.1313	0.0850	0.0197	0.1529	-0.2206	-0.1495	0.0658	-0.0592
29	-0.7911	-0.1394	0.0068	0.0117	0.0670	0.0453	-0.1298	-0.1367	0.0004	-0.0757
30	0.1135	0.0312	-0.7459	-0.1772	-0.1117	-0.0878	0.0870	-0.0238	-0.0694	-0.2485
31	0.5152	0.1468	-0.2864	-0.1791	-0.2623	-0.0541	0.0035	0.1822	-0.2620	-0.3152
32	0.2236	0.0698	-0.3773	-0.1665	0.0586	-0.1187	0.3520	0.0593	-0.2401	-0.3600

33	0.1173	0.2830	-0.2187	-0.4467	-0.1090	-0.0803	0.2634	0.1376	-0.0481	-0.3633
34	0.2165	0.1640	-0.2014	-0.1372	-0.4535	0.0577	0.1467	0.2435	-0.2806	-0.1486
35	-0.3605	0.2060	0.0911	0.1348	0.0526	0.4249	-0.2872	0.0756	0.1254	0.0255
36	-0.6628	-0.1412	0.0723	0.0615	0.0550	0.2327	-0.1140	-0.1514	0.1392	0.0110
37	-0.2733	-0.0818	0.0470	0.1828	0.0554	0.6315	-0.0778	-0.1945	0.1014	0.1240
38	-0.6348	-0.0188	0.0844	0.1197	0.0029	0.1995	-0.1583	0.0727	0.1091	-0.0151
39	-0.5192	0.0797	0.1389	0.2420	0.0114	0.3427	-0.0357	0.0929	0.1537	0.1701
40	-0.5339	-0.0449	0.0234	0.2981	0.2152	0.0457	0.1996	0.0402	0.1519	0.0565
41	-0.3115	-0.0289	-0.0493	0.5125	0.1487	0.1072	0.1387	0.1831	0.1150	-0.3117
42	-0.4300	-0.2213	0.0141	0.5499	0.1195	0.1279	0.0466	-0.1849	0.0880	0.0403
43	0.0753	0.1760	-0.7808	-0.0414	-0.1592	-0.1641	0.1041	0.0006	0.0078	-0.0546
44	0.1111	0.6649	-0.2183	-0.1194	-0.2855	0.0554	0.1328	0.0179	-0.1498	0.0009
45	0.3294	0.3509	-0.1422	-0.1948	-0.3945	-0.0270	0.0444	0.1189	-0.0755	-0.3006
46	0.0600	0.3365	-0.1822	-0.1529	-0.2772	-0.0539	0.5209	0.0794	0.0358	-0.0084
47	0.2041	0.3521	-0.1592	-0.1710	-0.4394	-0.0142	0.4347	0.2210	-0.1607	-0.0540
48	0.2213	0.3682	-0.1384	-0.0945	-0.4330	-0.0015	0.3747	0.2003	-0.1805	0.0326
49	0.1045	0.1910	-0.2656	-0.1347	-0.6439	0.0078	0.1056	0.1936	-0.0556	-0.0422
50	-0.0102	0.0560	-0.1009	-0.1130	-0.7378	0.0053	0.0369	0.0261	-0.0071	0.2066
51	0.0027	0.0388	-0.0189	0.0679	-0.5969	-0.1994	0.0404	-0.2279	-0.0179	-0.1098
52	0.2578	-0.0232	-0.1516	-0.1324	-0.3316	-0.0248	0.5351	0.0005	-0.1079	-0.1004

APPENDIX B

SPECIFIC INSTRUCTIONS AS ORIGINALLY DESIGNED FOR THE TREATMENT CONDITIONS

(a) SPECIFIC INSTRUCTIONS FOR CLASS I (T_1)

(These instructions are to be given in class, and woven into the instructors design of the "class activity." They are to be given orally.)

1. You will be expected to repeat each of the two within-term examinations at home. You may take up to four days before submitting this second attempt for scoring.

2. You will be free to make use of all resources, excluding instructors and fellow students. Your aim is to come out with all answers correct, working independently.

3. Part of your "class activity" score will be based on your performance in this examination repeat, and account will be taken of the gains you make in the number of correct responses.

4. (i) This part of the class activity is to count 10% of the instructor's grade, in other words it is worth 10 "points" out of a total of 100 "points" which make up the instructor's grade.

   (ii) Award 2 points to all subjects--for having carried out the exercise.

   (iii) Award the remaining 8 points according to the table below.

Initial Score On First Performance	Maximum Score On Second Performance	Maximum Gain	Points to be Awarded
80	80	0	8[1]
70_+	80	10	1 pt. for 1 gain
60_+	80	20	1 pt. for 2 gains
50_+	80	30	1 pt. for 3 gains
40_+	80	40[2]	
30_+	80	50	[2]1 pt. for 4 gains
20_+	80	60	
10_+	80	70	

[1]Note that the top scoring student apparently makes no gains but is awarded the total maximum points for "gains." He deserves it for maintaining his position in both performances. However, if he slips, his score on the second performance becomes the base and he is awarded points in the last category. For example, suppose second score is 68; the difference is 12 and his "points" 3. The instructor will be expected to comment on the practicality of this scheme after it had been used.

[2]Note that the rate is changed--to the favor of low scoring students on the first performance.

(b) SPECIFIC INSTRUCTIONS FOR CLASS 2 (T_2)

(These instructions are to be given in class and woven into the instructor's design of the "class activity." They are to be given orally.)

1. You will be expected to repeat each of the two within-term examinations at home. You will be free to make use of all resources excluding instructors and fellow students. Your aim is to come out with all answers correct, working independently.

2. You will also be expected to score and grade your two performances. Score, using your best judgments on what you feel are the correct answers. Evaluate your scores by assigning grades to yourself (0.....4.5), using some criteria you feel to be objective.

3. You may take up to four days before submitting your second performance for machine scoring.

4. Later when you receive the feedback, check your scoring and self evaluation and discuss the discrepancies with your instructor, until you are satisfied.

5. Finally prepare your Progress Chart and return it to your instructor for comments.

6. Part of your class activity score will be based on your performance in this exercise. Account will be taken both of the gains you make in the number of correct responses and in particular of the size of your mean discrepancies between your scorings and self evaluations and those of the instructor.

7. (i) This part of the class activity is to count 10% of the instructor's grade, in other words it is worth 10 points out of a total of 100 "points" which make up the instructors' grade.

   (ii) Award 2 "points" to all subjects--for having carried out the exercise.

   (iii) Award the remaining 8 points according to the mean discrepancy score as illustrated in the table on the following page:

TABLE OF POINTS TO BE AWARDED

Mean Discrepancy Score	Points to be Awarded
0 (Zero)	8
1 - 2	7
3 - 4	6
5 - 6	5
7 - 8	4
9 - 10	3
11 - 12	2
13 - 16	1
Above 16*	0 (Zero)

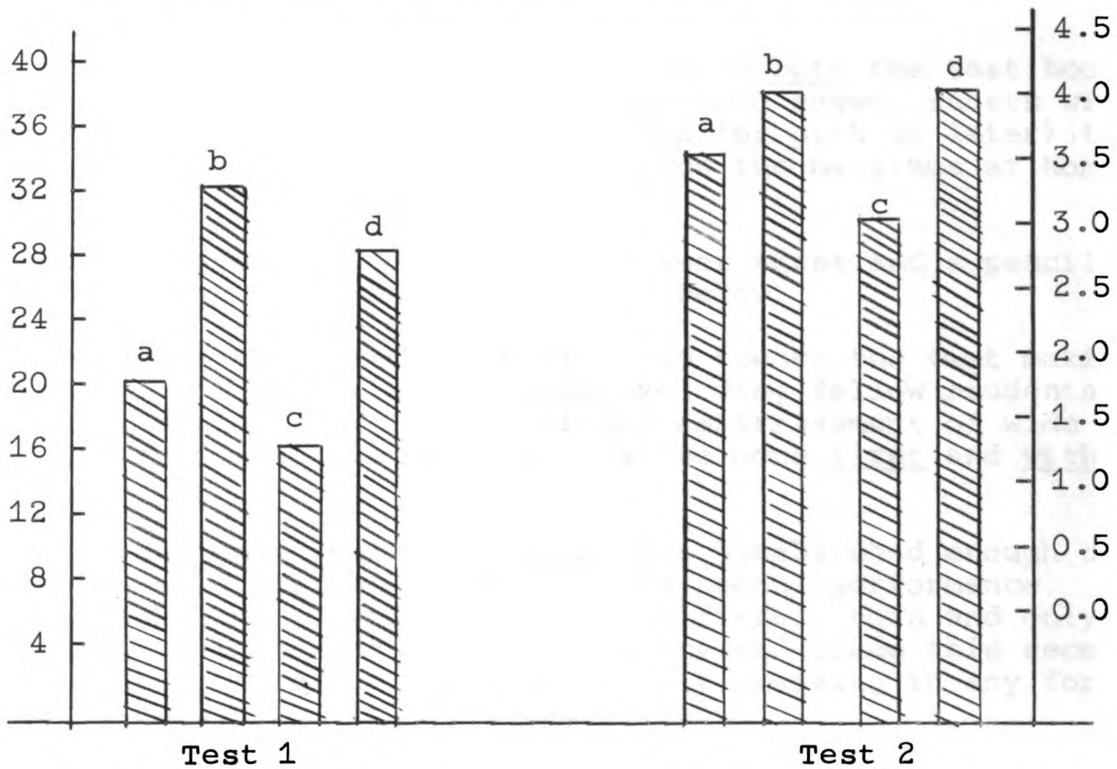
*16 (i.e., 20% of 80--the total maximum score) is the maximum discrepancy score that is to be rewarded.

The instructor will be expected to comment on the practicality of this scheme after it has been used.

8. The following is the Progress Chart to be introduced and explained to the student <u>after</u> the meeting to discuss discrepancies. The student will use <u>one</u> page of a Graph paper to prepare his chart as illustrated.

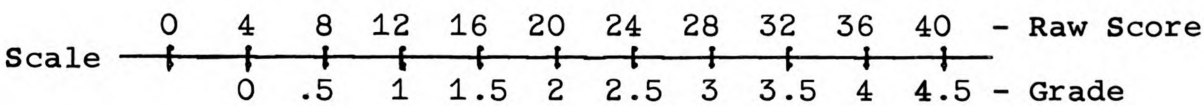
PROGRESS CHART

Aim: To Remove Discrepancies Between Evaluations



Key:

a = Self evaluation--in-class performance
b = Self evaluation--repeat
c = Instructor's evaluation in-class performance
d = Instructor's evaluation repeat



DETERMINATION OF MEAN DISCREPANCY SCORE

Item	Test 1	Test 2	Totals	Mean (N=4 pairs)
(a) minus (c)	4	4	8	$\frac{+12}{4} = +3$
(b) minus (d)	4	0	$\frac{4}{12}$	$= 3^*$
				*absolute value

(c) SUPPLEMENTARY INSTRUCTIONS TO CLASS 2 (T_2)

In administering your "treatment" the steps listed below should be followed closely:

1) Ask your students to

   a) write their names on their test booklets--to help them recover their copies

   b) mark their in-class performance on both the test booklet and the answer sheets provided; the answer sheets will be handed in but they will keep (or pick up later) their test booklets to score and grade the markings at home as described below.

2) Give to every student a spare answer sheet and a pencil for the repeat performance described below.

3) Emphasize that every student is to rework the test making use of all possible resources excluding fellow students and instructors. To prevent any embarrassment of wide discrepancies this exercise must be done first and with care.

4) When and only when the student has established enough confidence in his/her answers on the second performance (without any consideration of the first), then and only then should he/she proceed to score and grade this second repeat performance. Emphasize that guessing in any form will result in wide "discrepancies".

5) With the scoring and grading of his/her repeat performance as the "Key" the student then turns over to his marked test booklet to score and grade that performance also.

6) The student retains in his/her records his/her estimated score and grade. Then on a piece of paper, with his/her name on the paper, the following information is to be provided--ready to be handed in together with the repeat performance. Thus:

Name of Student_____

Test	Mid-term Test 2
In-class	Repeat
Estimated score	
Estimated grade	

This information will be used to check the accuracy of the graph.

7) In the following discussion class period the instructor collects the student's self-evaluations, and the repeat performance. Both must be collected before test results are to be made known in the times prescribed by the Course Coordinator.

8) When all the machine scores are returned to the student, the student prepares the graph (two copies of each) and returns them to the instructor.

9) The instructor then adds appropriate comments--the same on both graphs, one of which he/she keeps and the other returned to the student.

10) The instructor emphasizes that the graph is a Progress Chart--to give the student a visual image of his/her genuine progress. The graph also discourages guessing as it has been shown that this is the chief factor in wide "discrepancies".

APPENDIX C

MEAN PERCENT OF RESPONDENTS CHOOSING OPTION ON THE FACTOR SUB-SCALES

	1			2			3			4			5		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
EP															
Learning Function	26	27	28	34	32	33	23	27	24	14	11	12	3	2	2
Motivating Function	24	34	28	34	36	35	22	12	21	16	15	13	2.5	1.5	1.5
GP															
Learning Function	27	28	28	28	30	29	27	24	25	13	15	14	4	3	4
Motivating Function	11	14	16	28	23	24	30	33	27	27	25	29	4	5	4
EN															
Dysfunction	7	6	6	31	36	28	34	33	35	20	26	22	7	8	8
Pressure-Anxiety	4	6	3	22	12	17	22	23	24	30	27	32	21	24	21
GN															
Dysfunction	6	8	7	21	24	18	25	23	33	31	26	25	17	18	16
Pressure-Anxiety	2	4	2	10	13	14	22	15	19	37	39	33	28	28	32