THE EXPERIMENTAL EFFECT OF GRADES ASSIGNED TO A SINGLE TASK ON SUBSEQUENT ACADEMIC PERFORMANCE

THESIS FOR THE DEGREE OF PH. D.

MICHIGAN STATE UNIVERSITY

DAVID MORGAN NOLAN

1964

THESIS

This is to certify that the

thesis entitled

THE EXPERIMENTAL EFFECT OF GRADES ASSIGNED

TO A SINGLE TASK ON SUBSEQUENT ACADEMIC PERFORMANCE

presented by

David Morgan Nolan

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Education

William M. Vaughan
Major professor

Date June 22, 1964

THE EXPERIMENTAL EFFECT OF GRADES ASSIGNED TO A SINGLE TASK ON SUBSEQUENT ACADEMIC PERFORMANCE

By

David Morgan Nolan

A THESIS

Submitted to the School for Advanced Graduate Studies
of Michigan State University of Agriculture and
Applied Science in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Administrative and Educational Services

1964

TABLE OF CONTENTS

LIST OF TABLES

THE EXPERIMENTAL EFFECT OF GRADES ASSIGNED TO A SINGLE TASK ON SUBSEQUENT ACADEMIC PERFORMANCE

By

David Morgan Nolan

AN ABSTRACT

Submitted to the School for Advanced Graduate Studies
of Michigan State University of Agriculture and
Applied Science in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Administrative and Educational Services

1964

Approved_____

David M. Nolan

ABSTRACT

It was the purpose of the study to test the theory that grades assigned to students' work function as a reinforcement and thus effect performances subsequent to the receipt of the grades.

The experiment was conducted with students enrolled in an educational psychology course (F. E. 200) at Michigan State University. In two sections of the course there were 225 undergraduate students (99 in the first section and 126 in the second) for whom all information was available pertinent to the experiment. There were 68 males and 157 females in the experimental population.

A treatment-by-levels with one replication design was used to test the hypotheses. The students were divided into three levels based on their grade point averages (Hi, Mid, and Lo). The treatment variable consisted of the random assignment within each level of grades of A, C, or F to an essay examination taken by the students. The criterion variable consisted of the scores attained on an objective test covering the next assignment and administered at the next class meeting. An effort was made to include a second control variable based on the students' attitudes toward work persistence. However, the instrument intended to measure this attitude failed to function at a level high enough to warrant

the use of scores obtained with it.

Evaluation of the results was accomplished by means of an analysis of variance to determine what, if any, differential effect occurred as a result of the various treatments within levels. The analysis was done for the total population and separately by sex. In no case was the main effect (treatment alone) found to produce significant results. Likewise, in all cases the simple effect (interaction of treatment by levels) failed to show significant results.

The limitations of the study were:

1. Failure of the work persistence attitude scale to provide an effective control.

2. Ethical restrictions limiting the treatment variable to a single application.

3. The fact that all subjects were at least sophomores in college and thus less likely to show the effect of a single treatment.

4. All subjects were enrolled in a required course leading to a teaching certificate.

5. It was possible that the treatment was not internalized by the students.

It was concluded that within the limitations of the study there was no evidence to indicate that grades assigned to one student performance have an effect on a subsequent performance.

CHAPTER I

DEFINITION OF THE PROBLEM

A frequent bit of advice given to new teachers is to grade low. The reason generally given has to do with student motivation. The tone of the advice and the advice itself lead to the idea that grades are rewards and that by arbitrary manipulation of these rewards the teacher may gain the desired academic performance.

A low grading procedure could have a number of effects. For some students it might result in harder work in order to achieve the grade to which they aspired. For some, who may not be particularly grade conscious, it might not have any effect whatsoever. For others it might serve as a reinforcement of an attitude of academic inferiority and result in a lower classroom performance. An easy grading policy might be expected to have similar effects. Some students may be encouraged, and, as a result, strive for further rewards. Some may not be affected, and some might find themselves in a situation of a cheap reward and, therefore, a resulting lower performance.

The present study was an investigation of the effect grades received on one assignment had on students' performance on an assignment immediately following.

Theoretical Statement and Assumptions

In theoretical terms, the problem and the study may be thought of as dealing with situations of positive and negative reinforcement. Hilgard, in his book Theories of Learning[1], provides a model for such an approach in his summary and attempt to mediate the various learning theories he has reviewed. It may be assumed that students' classroom behavior is motivated, an essential in a learning situation. Despite the desirability expressed by Hilgard for intrinsic motivation, the extrinsic goal of a good grade may be considered an adequate goal for generating sufficient motivation to make a "provisional try... The theory supposes that a provisional behavior route is kept in suspension until its consequences change its provisional status; if it is confirmed it is an appropriate path of action to be followed."[2] Thus, students who have had extensive experience in receiving high grades are positively reinforced when given another high grade and would be expected to continue striving for further rewards. A negative reinforcement with the same kind of student would be the assignment of a low grade which would be expected to result in a change of behavior from that which had previously produced high grades.. Other combinations are possible which might produce a number of reactions:

1. Low-grade-point-average students given low grades would be positively reinforced and thus expected to perform at their customary low level.

---

1. Ernest R. Hilgard, "Reinforcement, Provisional Try, and Feedback," Theories of Learning, Appleton-Century-Crofts, Inc., New York, pp. 469-472.

2. Ibid.

2. Low-grade-point-average students given high grades would be negatively reinforced and a change in the direction of the negative reinforcement might be expected.

3. Lesser reinforcement (positive or negative) would be expected with middle-grade-point-average students when given high or low grades and, therefore, would be expected to result in less change.

4. High and low-grade-point-average students when given middle grades would not be receiving strong reinforcement in either direction and might be expected to react with less change than the extremes.

Thus, it becomes evident that the problem is compounded by the past academic achievement of the students. That is, a strong positive reinforcement to one student may be strongly negative to another and mild to still another, all depending on past experience in receiving grades.

A second delimiting factor is likely to exist within the above theoretical scheme. Individual differences among students within the different grade levels would result in students being more or less resistent to the reinforcement. For example, although there may be an average reaction among high-grade-point-average students to a negative reinforcement, the differing attitudes of the individual students toward their work and the rewards accruing from it, might be expected to cause some to react more strongly than others to the negative reinforcement.

Basically then, the problem is to empirically investigate, within the controls of grade level and attitude toward work, the short-term effects of grading hard or easy on students' academic behavior as measured by an achievement test.

Hypothesis

From the theoretical statement above, the following brief hypothesis may be made. It is hypothesized that: The academic performance of students subsequent to arbitrary positive or negative reinforcement in the form of assigned grades will be:

1. influenced by the reinforcement and
2. influenced in the direction of the reinforcement and by the students' attitudes toward persistence in their work.

The reason for the two hypotheses, one nondirectional and the other directional, is the lack of evidence from past studies as to what effects might be expected. For instance, there is a school of thought[3] which claims that the assignment of certain grades results in complacence and an easing in effort. Results of studies on the subject are not clear as to what direction an effect might take.

Overview of Study

Chapter Two is devoted to a review of other research which has been done in the general area of the effects of various kinds of evaluation on further performance.

---

3. P. J. Fay, "The Effect of the Knowledge of Marks on the Subsequent Achievement of College Students," Journal of Educational Psychology, 28: pp. 548-554; 1937.

In Chapter Three the design of the study is explained. It includes descriptions of the possible errors, assumptions, and the null hypotheses. In addition, the experimental procedures are described.

The fourth chapter contains descriptions and characteristics of the measurement instruments used in the study.

Chapter Five constitutes the actual analysis of the data with appropriate graphic presentations.

In Chapter Six, a summary is presented with an emphasis on the conclusions and implications of the study. This chapter also includes the observed limitations of the methods used.

CHAPTER II

REVIEW OF PERTINENT RESEARCH

There is general agreement in the literature that research is needed on the subject of the effects of grades on student performance. In the 1950 edition of the Encyclopedia of Educational Research, C. W. Odell[1], editing the chapter on Marks and Marking Systems, wrote that there had been extensive attention given to grades and grading practices, but that there were few studies giving some semblance of objective data of the results of attempts to improve marking systems. He specifically observed that "the question of the effects of marks upon those to whom they are given is very important and merits much study. Much opinion has been expressed, but few facts are available."

Ten years later Max Wingo[2] stated in the 1960 edition of the Encyclopedia of Educational Research that "more specific information is needed in this area, particularly with respect to such questions as these: ...What are the effects of evaluation on student attitudes, understandings, and behavior patterns?"

---

1. C. W. Odell, "Marks and Marking Systems", Encyclopedia of Educational Research, New York, McMillan, 50: pp. 711-717.

2. G. Max Wingo, "Methods of Teaching", Encyclopedia of Educational Research, New York, McMillan, 60: pp. 848-859.

In order that a consistent approach to the writing in the field under perusal may be assured, the outline for evaluating research in psychology and education presented by Farquhar and Krumboltz[3] has been employed wherever appropriate. The outline covers five basic parts of any research presentation: the problem, the design, the procedure, the analysis and the interpretation.

The original intention was to review the literature for the past ten years, but truly pertinent research was as scarce as Max Wingo[4] had warned. It was therefore, decided to include what could be found prior to 1950, despite Odell's[5] similar warning concerning research prior to that year. In some of the older studies it would not be appropriate to apply the rigorous evaluation advocated by Farquhar and Krumboltz[6], and in such cases, modifications are made.

On the whole, Odell[7] and Wingo[8] were found to be correct in their evaluation of research done in the field of the effects of marks on the recipients. Odell commented that there were few studies with "even semi-objective data."

---

3. William W. Farquhar and John D. Krumboltz, "A Check List for Evaluating Experimental Research in Psychology and Education," Journal of Educational Research, 52: pp. 353-354, 1959.

4. Wingo, op. cit., p. 848.

5. Odell, op. cit., p. 711.

6. Farquhar and Krumboltz, op. cit., p.353.

7. Odell, op. cit., p. 717.

8. Wingo, op. cit., p. 859.

Most of the studies reviewed are simply not worth reporting. An example of a typical piece of research whosetitle tempts a plunge into the dusty archives of the '20s is "Pupil Reaction to School Reports" by W. A. Barton.[9] There was no statement of the problem at all. The design, procedure and analysis consisted of asking an undisclosed number of high school students questions about marks and reporting their responses in percentages. It may well be that Barton was forced to leave out much of his study in order to satisfy publication requirements. The following are some of his results:

1. 60% said marks helped in planning.

2. 56% said they were more interested in a passing mark than the subject.

3. 67% preferred percentage marks.

4. 21% preferred letter marks.

5. 6% preferred numbers.

6. 77% boys and 53% girls said low marks make them work harder.

7. 43% boys and 48% girls said low marks make them feel they are failures.

8. 48% boys and 54% girls said high marks make them work harder.

---

9. W. A. Barton, Jr., "Pupil Reaction to School Reports," The School Review, 33:771-780; 1925, 34:42-53, 1926.

Because there is so much missing from the report of this study, it is difficult to draw any conclusion other than that there were mixed feelings among high school students toward marks in 1926.

One year prior to Barton's questioning, E. B. Hurlock[10] conducted an experiment in Harrisburg, Pennsylvania, that deserves credit as one of the most diabolical ever perpetrated on a group of elementary school children. The problem was clearly stated: In the classroom, do children who constantly receive praise for their work show more improvement from day to day than do the children who are reproved or completely ignored. The hypothesis growing out of the problem is evident in the design and procedure of the experiment. Hurlock used all the students in grades four and six in an elementary school. There were 48 boys and 106 girls. Her criteria of improvement were the scores from the Curtis Research Tests in Arithmetic, Addition Form. There were five tests given on five consecutive days to the whole group. Each test contained thirty items of equal difficulty. All pupils participating were given fifteen minutes a day for practice. On the basis of the results of the first test, the pupils were assigned to four equivalent groups. Group C or control were isolated from the others while the treatment was applied. Group

---

10. E. B. Hurlock, "An Evaluation of certain incentives Used in School Work," Journal of Educational Psychology, 16: pp. 145-159; 1925.

P or praised were individually called to the front of their class and praised and encouraged. Group R or reproved were individually called to the front of their class, told how stupid they were and generally criticized for their performance. Group I or ignored were just left sitting in the class while the P's were praised and the R's punished. The treatments were randomly assigned but there was no replication and levels of significance were not prescribed.

The results of the ordeal were as follows:

P: showed steady improvement each day for the full five days.

R: showed just as much improvement as P the first two days, then dropped slightly and leveled off the last two days.

I: improved at first then dropped and leveled off below R.

C: did not show as much improvement as any of the others and were well below the others at the end of the five days.

The statement was made that the difference between P and C at the end of the five days was the only one large enough to be statistically reliable, but no figures were reported. A breakdown by sex and age showed slight differences, but the significance was not reported. Hurlock reported that the inferior pupils (that is, those with the lowest scores on the first test)

had the greatest gains under both praise and reproof. He did not report the significance of the differences. It appears likely, however, that the difference between the inferior and the superior pupils was a function of the limits of the test and the effect of regression toward the mean, rather than a result of the experiment. The mean score for P on the fifth day was 20.22 with a standard deviation of 7.68 which means that one S. D. above the mean was a score of about 28, or just 2 points below the maximum possible. It would appear that the superior in at least the P group had topped out on the test, thus making any comparisons of improvement between ability classifications invalid.

Although a number of conclusions and generalizations were made, none of them are fully supported by the data except that praise showed better results than nothing. That is, P resulted in greater gains than C. No statement was made of how the experimental effect was obliterated.

In 1937, P. J. Fay[11] made an effort to measure the effects of knowledge of marks on achievement in a beginning psychology class.

The problem was clearly stated as an investigation to determine the effects on achievement of a closed marking system (students were told only that they were passing) compared to an

---

11. Fay, op. cit, pp. 548-554.

open system (students were given letter grades, A, B, C+, and C, based on a test administered at the end of each four weeks of the semester). The open system was designated experimental (E) and the closed system, control (C). Students attaining D or F grades on the first test had to be reported and were thus forced out of the experiment. Greater achievement with the open system as measured by gain between the first and final tests was hypothesized on the theory that knowledge of progress would result in greater incentive.

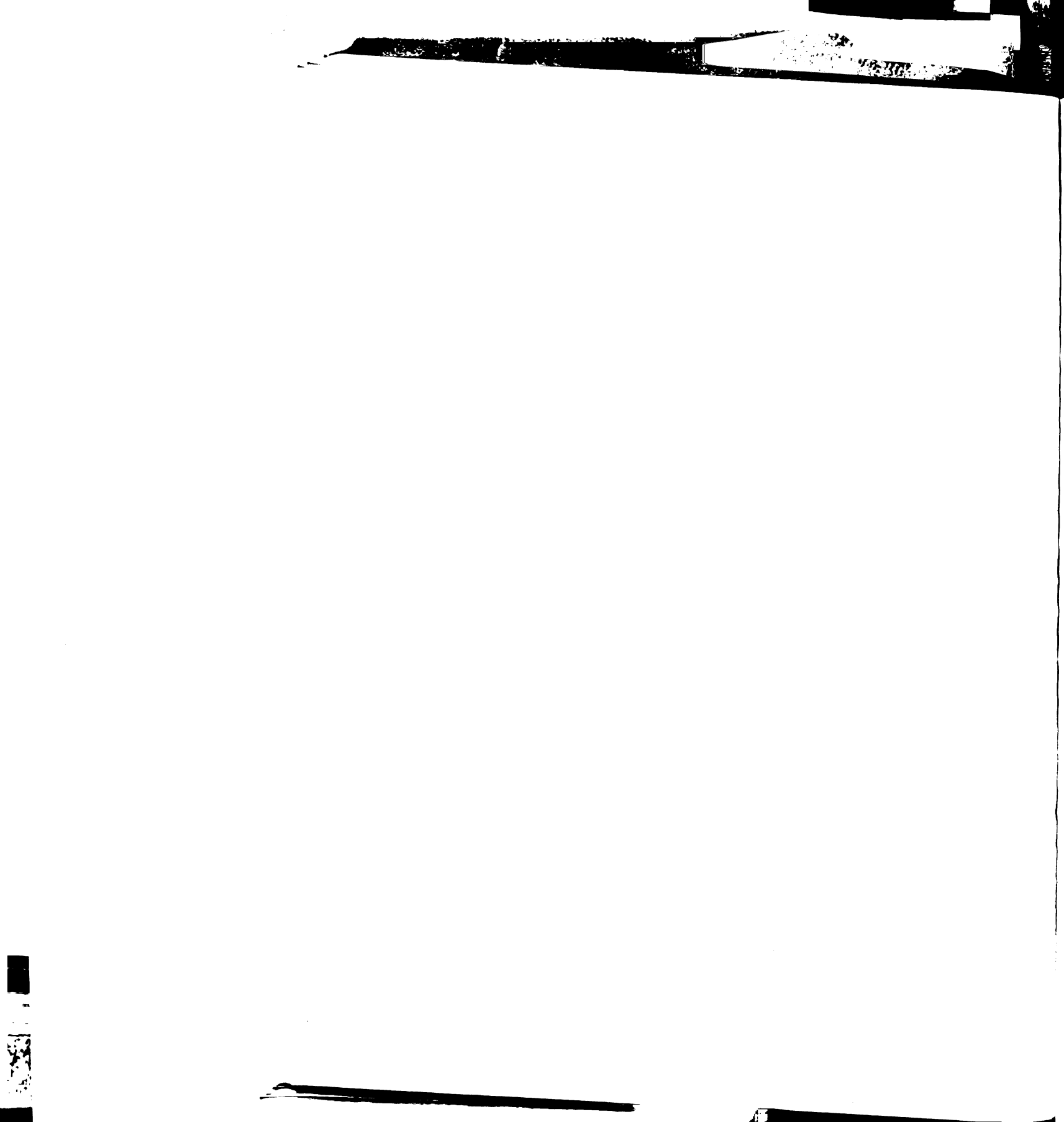
One hundred and ninety-six juniors and sophomores in a beginning psychology class were divided into control and experimental groups on the basis of ACE Psychological Examination scores and the scores on the first monthly test. All students attended the same lecture session, but were divided into eight discussion sections. Two instructors shared the lecture responsibilities and divided the discussion sections equally, each taking two experimental and two control sections. At the end of each four weeks an objective test of 125 to 150 items was administered. E students were given their grades on each of the tests. C students were told that they were either passing or failing. The final exam consisted of 400 items. All test scores were converted to a standard scale with a range of 0 to 100, a mean of 50 and a standard deviation of 14.

The statistical design consisted of correlations between test scores and gains or losses. Changes in a group's position were reported as probability statements. Thus, for experimental group A (those receiving an A on the first test), which increased 1.15 points between the first test and the final examination, "there are 62 chances in 100 that this is a true increase." At the time this experiment was conducted the statistical techniques available were rather limited. Over criticism of the technique employed is therefore, not called for.

Fay made no attempt to conduct the experiment with a random sample of the college population although the conclusions he drew were not qualified by a statement of the limitations of the population. The control group, as previously described, was adequate. There was no replication, nor were levels of significance prescribed.

An exact replication would not be possible, although an independent investigator could approximate the experiment closely enough to check the findings.

The size of the sample was adequately described, but other than the statement that they were sophomores and juniors, nothing further was said to describe it. No data on the ability level of the students was reported, although Fay had ACE scores on them all. The experiment was conducted in such a way that



differences could be attributed to experimental effect.

The number of items on the tests, the ranges, standard errors, standard deviations and means were reported, but nothing was stated which established the tests as valid measurements of the material being studied in the course. The standard error of the tests indicated a high enough reliability. There was no mention of the appropriateness of the statistics employed.

As has already been mentioned, Fay did not limit his generalizations and conclusions to the population on which he conducted his experiment.

Three basic conclusions were drawn from the experiment. All are questionable on the basis of the limitations already mentioned. However, quite specific factors serve to invalidate each of them, as well.

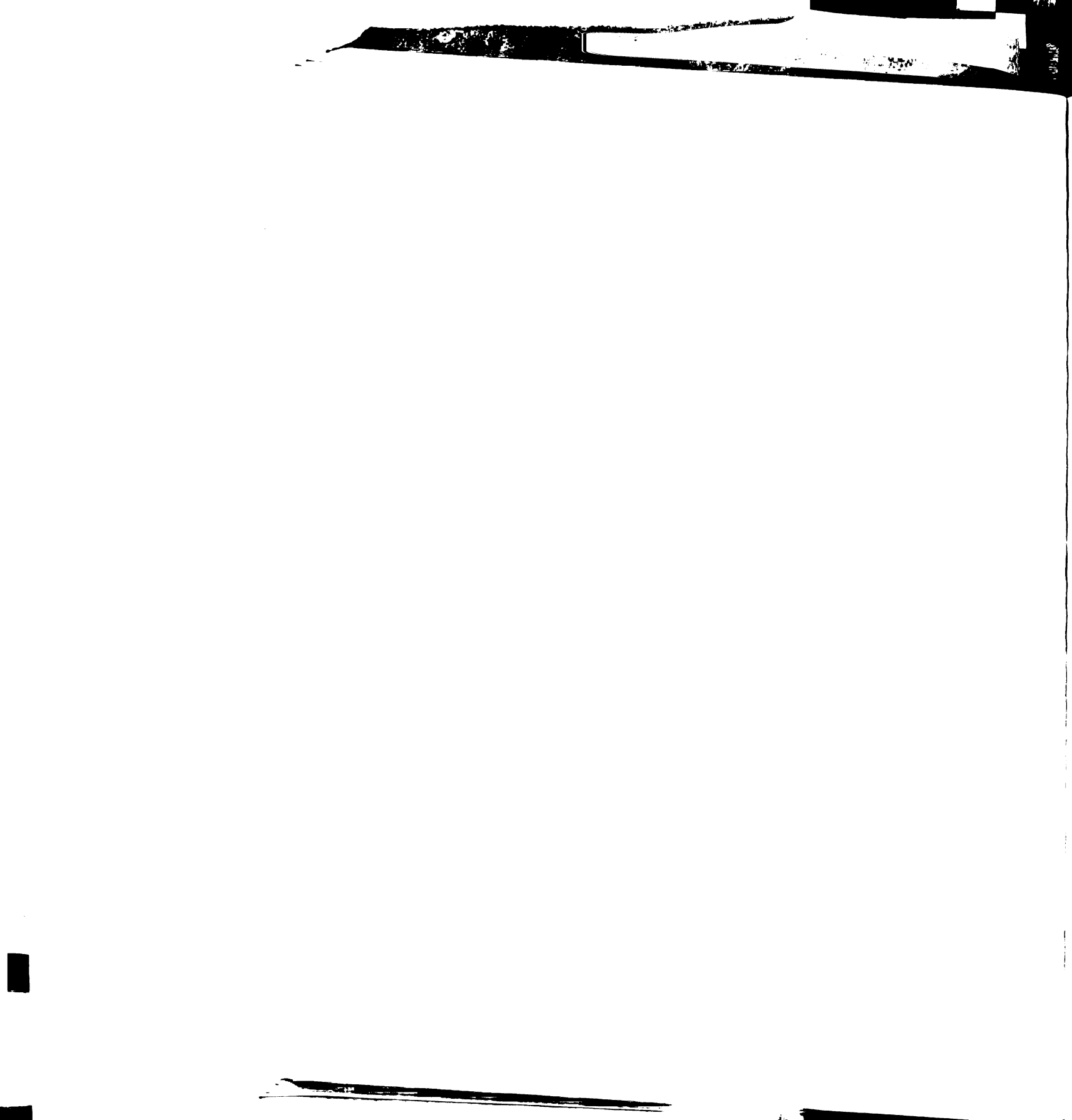
The first conclusion which was not covered by a hypothesis stated that if A and C students know their grades, they work harder to retain or improve them, and that when B and C+ students know their grades, they are satisfied, become complacent and, therefore, slip back. This conclusion was based for the experimental A group on a rise of 1.15 points on the standard scale between the first test and the final examination. The experimental C group had an even smaller rise and the B and C+ students had equivalent drops. The conversion to standard

scores introduced an ipsative factor making any comparison of gains or losses between tests of limited value and certainly not warranting the conclusion drawn. The nature of the standard scale is such that if one group rises, some other group or groups must fall. If the A and C students did much better on the final test relative to the first test and the B and C+ students held their raw score positions relative to the first test, the conversion to the standard scale would make it appear that the B and C+ students had dropped. It is quite possible that the B and C+ students improved relative to the first test, but not as much as the A and C students, again resulting in an apparent drop on the standard scale for the B and C+ students. Actually, the only conclusion possible from the data presented is that one or more of the groups moved in comparison to the others. However, the size of the changes rule out even that possibility.

The second conclusion was offered in support of keeping students informed of their grades. It stated that when students know their grades, they are more apt to retain their relative positions (based on the first test) on the final examinations than when they do not know their grades. Even if such a thing could be considered pedagogically desirable, it is based on the following correlations between the first test and the final examination:

Experimental: +.71

Control: +.60



Because no data is provided showing the reliability or the significance of the difference, it is impossible to accept the conclusion.

"It is better to keep students informed of their grades" was the third conclusion. It is based on correlations of the gains or losses between the first test and the final examination and the ACE percentile rank. For the experimental group (who knew their grades), the correlation was negative .05. For the control group, it was positive .43. Such a difference certainly appears significant even though no statement was made regarding its significance.

Fay accounted for the difference by saying that the lower IQ's in the experimental group gained as much as the higher IQ's because of the added incentive of knowing their grades, thus resulting in a low correlation. In the control group, without the extra incentive, the low IQ's did not gain, while the brighter students did because they were bright. Even without pointing out the obvious alternative to this conclusion, it may be discounted from the data. A table showing the number of gainers by ACE percentiles indicates that there were the same number who gained in the experimental and control groups below the 40th percentile. However, the control group had more gains between the 60th and 90th percentiles. The control group's greater gains

between 60 and 90 account for the higher correlation and, without attempting an explanation of the reason for C's gains, the temptation is for a conclusion the exact opposite of Fay's. The temptation must be resisted, however, because of the previously explained weaknesses in the experiment. Fay's experiment and conclusions have been used as a defense of the use of grades. Such use appears completely unwarranted on the basis of the published account of the experiment.

An experiment conducted in 1941, in Texas, gives some evidence that grades do not do what teachers think. Portwood[12] reported the results of the elimination of a failing grade in a Negro high school in San Antonio. The problem was ill defined, the design was lacking, the procedures were haphazard, the analysis was almost nonexistent and the interpretation left much to be desired. About all that can be said for the experiment was that it should have given the experimentor a hunch, and he should have followed it up. There is no evidence that he did. The hunch is the only reason for including Portwood's experiment.

At the beginning of the school year the 650 students were informed that the F grade had been eliminated and no student need worry about failing. The student attitude at the beginning was typified by the statement: "Why work, if I can't flunk?"

---

12. T. B. Portwood, "Success Does It-When Pupils Can't Fail, They Don't,: Nations Schools, 28: pp. 60-62; 1941.

The teachers had a rather similar reaction: "Pupils won't come to school if they know they can get by," and "How can I force students to work?" Portwood stated that attendance did not drop, that the teachers reported increased interest on the part of students, that the technique resulted in better teacher planning and better guidance. How he arrived at any of these conclusions was not related. If proper controls and evaluation criteria had been included in the experiment, the results could have been significant, because the F for failure must have a special value in a low socioeconomic group such as an all-Negro high school in San Antonio. As it is, there is only the hunch that the elimination of the F may have had an effect on both the students' and the teachers' performance.

A number of articles expressing personal and collected opinions about grades and their effects appear throughout the years. Typical was an article by Ruth Little[13] which consisted of a large number of statements from teachers about grades. The title was a question: "Whither Grading" and the apparent answer was: We do not know.

In 1961 Bostrom, Vlandis and Rosenbaum[14] reported a

---

13. Ruth C. Little, "Whither Grading", National Education Association Journal, 36: pp. 12-13; 1947.

14. Bostrom, R. N., "Grades As Reinforcing Contingencies and Attitude Change," Journal of Educational Psychology, April, 1961, V. 52, No. 2.

study in which they used grades as reinforcements in the changing of attitudes. Bostrom commented on the lack of research related to the study with the statement that, "Little, if any, information is available about the effects of grades on the students to whom they are given."

The problem was clearly stated as an examination of the effect of differential assignment of grades to essays, on subjects about which the students would be likely to have preformed attitudes. The underlying hypothesis was that grades will act as a reinforcing contingency and thus affect behavior. Specifically, good grades should result in repetition of previous attitudes while poor grades will decrease the likelihood of repetition of previous attitudes. It should be noted that while the hypothesis appears to be only a restatement of what any and all teachers hope to accomplish through the assignment of grades, it deals not with the quality of the work to which a grade is assigned, but with the attitude expressed by a student in the work.

The hypothesis was based on the theory of Doob[15] that an attitude is an anticipatory response which mediates overt behavior, but which in turn is derived from the reinforcement of overt behavior.

---

15. L. W. Doob, "The Behavior of Attitudes," Psychological Review, 1947, 54: 135-156.

The study was conducted by assessing the attitudes of 228 students in communication skills classes with a questionnaire on four attitude-influenced subjects. Because the results for one subject were highly skewed and for another unreliable, the study was conducted with only two of the subjects measured: legalized gambling and socialized medicine. The position assigned was opposite to that determined by the scale and the topic assigned was the one on which the particular student showed the strongest initial position. The essays were written in class within a time limit of 30 minutes. The students were promised grades on the following day. Grades were assigned randomly; one third was given A, one third was given D, and one third was not given a grade. Students in the no-grade group were told that there had not been time to assign them grades. Immediately after the essays were returned, the attitude questionnaire was re-administered.

Change scores on the attitude scale were determined by subtracting post-test scores from pre-test scores. Because no information was provided regarding the reliability of the scale scores or the change scores, it is not possible to comment on the level of significance used as a criterion of true differences between groups. Otherwise, the statistical design was appropriate. The mean change for each group (A, D, and No-Grade) was computed and a *t* test recommended by Edwards[16] for heterogeneous variance

16. A. L. Edwards, Experimental Design in Psychological Research, (Revised Edition), New York, Rhinehart, 1960.

was employed to check the significance of the differences between the groups. The significance of the difference between group A and D was .01, between A and No grade .05, and between D and no grade less than .10. The population from which the sample was drawn was not specified and the selection of the sample was not described. The control group consisted of the No-grade group and was selected at random from the sample. Treatments were assigned at random. The level of significance for rejection of the null hypothesis was not stated although the conclusion did make the rejection.

Because the attitude scales were not adequately described, it would be impossible to do an exact replication of the study. Otherwise, the procedures were sufficiently well described to permit replication. Although the size of the sample was given, other characteristics were ignored. The treatments were randomized so that differences could be attributed to experimental effects.

The main weakness of Bostrom's study is the criterion of evaluation. The entire study hinges on the reliability and validity of the attitude scales. The fact that one scale was discarded because it was unreliable indicates that some reliability information was acquired on the scales, but none was reported. Nothing at all was said about the validity of the scales.

The test of the hypothesis was based on the change scores from the attitude scales. Without knowing the reliability of these

scores, it cannot be concluded that the statistical assumptions for the test were met.

The conclusion that good grades were demonstrated to serve a reinforcing role in contrast to the effects of poor or no grades was consistent with the obtained results. Qualifications were made concerning the generalization as far as the limits of the experiment and the sample were concerned.

Despite the severe limits imposed by the inadequate treatment of the criterion measure, there can be little doubt that Bostrom et al demonstrated a definite change in attitude resulting from an assigned grade.

If a grade on a single essay can be demonstrated to have a causitive effect on an attitude concerning socialized medicine, it seems reasonable to conclude that grades could also have an effect on attitudes concerning learning, school and specific subjects in school.

Another demonstration of the effect the results from a single task had on student behavior was accomplished by Page[17] in his study of the effects of teacher comments on subsequent student performance. This study was not aimed at measuring the effect of grades or the manipulation of grades as such, rather at measuring the effect of written comments about students' performance. It showed that students' work may be affected by

---

17. Page, Ellis Batten, "Teacher Comments and Student Performance: A Seventy-Four Classroom Experiment in School Motivation," Journal of Educational Psychology, August, 1958, V. 49, No. 4.

by the teacher's reaction to previous work.

Page[18] stated the problem clearly: What effects have teachers' written comments on students' papers on the students' future performance? Directional hypotheses were not stated and the non-directional hypothesis was implied by the design.

Random sampling procedures were followed throughout: The selection of seventy-four secondary school teachers, the teachers' selection of one each of their classes, and the assignment of students from that class to one of three treatment groups. The procedure consisted of grading an objective test in the usual fashion by assigning a letter grade to each test peper. The papers were matched in groups of threes according to the grade received and assigned a treatment of:

1. No comment (N), just the letter grade;
2. A specified comment (S), according to grade received;
3. A free comment (F), whatever the teacher chose to write.

The scores on the next objective test administered by the teacher were reported to the experimenter and became the criterion measure. The ranking of students within levels (matched groups of three) on the second test was the data used in the analysis of variance.

The results showed differences in performance on the criterion tests as follows: F significantly higher than N, F higher

---

18. Page, op. cit.

than S but not significantly, and S significantly higher than N. Page concluded that teachers' written comments "have a measurable and potent effect upon student effort, or attention, or attitude, or whatever it is which causes learning to improve."[19]

There is no question that Page's data showed significant differences and that the study was done with adequate procedures and analytical techniques. However, his conclusion is not the only one possible from the results. It would have been quite possible to conclude that those students who did not receive written comments, and who probably knew that many of their fellow students had received them, felt neglected by their teachers and as a result did not learn as much, or try as hard, or "whatever". The weakness in this study then, is the lack of control groups covering the communication between treatment groups.

The study, nevertheless, demonstrated that it is possible for some kinds of evaluative reactions (grade, comment, or lack of comment) to influence a student's academic performance over a short period of time.

Summary

In summary, it can be said that despite the recognition of the need for research on the problem of the effects of grades on student behavior and attitudes, little of significance has been done. The work of Bostrom, Vlandis and Rosenbaum[20], in 1961,

---

19. Page, op. cit.

20. Bostrom, op. cit.

indicates an effect on an attitude from a grade. Page's[21] study showed some direct effect on students' academic behavior. Although there were some efforts to measure the effects of grades on student performance, the limitations of the studies rendered the results inconclusive.

21. Page, op. cit.

CHAPTER III

THE DESIGN OF THE STUDY

The Treatment by Levels Design

There are a number of statistical designs available for the analysis of a study attempting to determine the effects of various treatments on a population. Following the recommendation of Lindquist[1], it was determined to use the treatment-by-levels design with an analysis of variance. The major purpose of this design is to increase the precision of the treatment comparisons by matching the treatment groups with reference to a control variable related to the criterion variable.

The Control Variables

Because of the likelihood that the criterion variable (scores on an objective test) would be positively correlated with the academic ability of the students, the experimental population was divided into three levels based on grade-point-average. It was this division into levels that increased the precision of the experiment. A second control variable about which the presumption of correlation with the criterion is not possible

---

1. E. F. Lindquist, "Treatment X Levels Designs," Design and Analysis of Experiments in Psychology and Education, Houghton Mifflin, Boston, pp. 121-149.

is the scores obtained on a ten-item Attitude Scale. An Analysis of this scale was required before it could be considered a true control variable.

The Treatment Variable

In view of the objectives of the study, it was necessary to use as a treatment some kind of an arbitrary manipulation of the grades received by the students. Thus, the treatment variable consisted of the arbitrary grades (A, C or F) assigned at random within levels to essay tests completed by all the students.

Criterion Variable

Again, the objectives of the study dictated the device used to determine the effects of the treatments. This device was some measure of academic performance obtained following the treatment. Thus, the scores on an objective test administered three days after the application of the treatment constitute the criterion variable. Following Anastasi's[2] recommendation, the Kuder-Richardson Formula 21 was used to gain a measure of the objective test's equivalence and homogeneity.

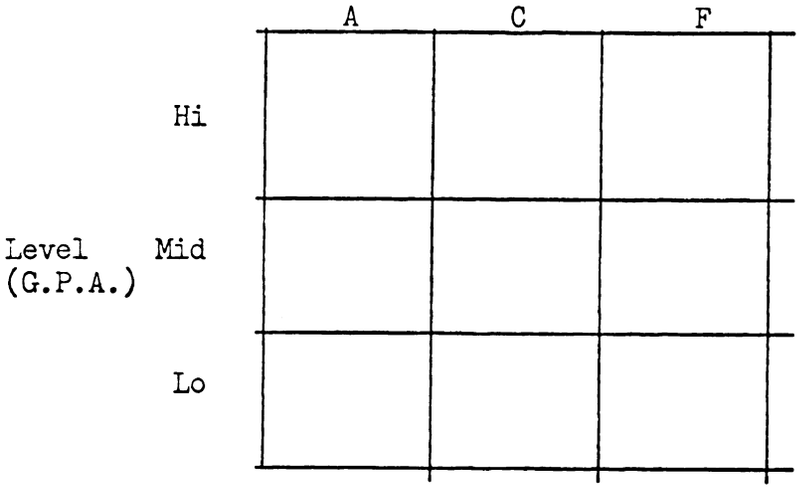
Diagrammatic Plan of Design

It is possible to show the treatment-by-levels design

---

2. Anne Anastasi, "Test Reliability," Psychological Testing, MacMillan, New York, p. 110.

described above graphically:



The Control of Error Variables

Following Lindquist's[3] approach to analysis ofvariance there are three basic types of error possible in the design selected for this experiment:

> Type S: Result from the assignment of subjects to the treatment groups. Even though the assignment is done at random, there is still a chance that in any single replication of an experiment, students with greater ability, motivation, and/or interest will be assigned to one particular treatment group

3. Lindquist, op. cit, pp. 8-11.

rather than being equally distributed throughout. The use of the levels design minimizes the likelihood of the Type S error. In addition, the available test of significance "is valid so far as type S errors are concerned."[4]

Type G: Result from differences in the application of the same treatment to different groups. Systematic differences in the criterion means can be caused by differences in the application of the treatment. For example, even though treatment A may, in truth, be a more effective procedure than treatment F, if A is administered poorly, the criterion measure may not show any difference or may even show F to be superior. The design of this study is such that Type G. errors are held to a minimum. The treatments were administered to the individual subjects at all levels and within all treatment groups at the same time and in the symbolic form of an A, a C or an F appearing on the essay test. The randomization of treatments independently within levels further reduced the

4. Lindquist, op. cit., p. 137

possibility of a systematic Type G error.

Type R: Result from variations in the treatment effects due to characteristics of the subgroups or replication samples. It is possible that sub-populations exist which, because of certain characteristics, are more or less affected by the treatment than is the population as a whole or other sub-populations. Because this experiment used the entire population and assigned the treatments at random within ability levels, the danger of Type R errors were considerably reduced within the study. It may be that a replication with a group of students at a different year in their academic experience or from a different institution or within a different department would reveal the existence of Type R errors if generalizations were made from the results of this experiment to students as a whole.

In addition to the above design elements intended to ...e the possibility of error, a replication was done, with the only difference being that the second experiment was conducted with students whose class meeting occurred one hour later in the day. This replication provides a cross-validation of the

results obtained with the first group and thus further reduces the likelihood of systematic sampling errors being interpreted as treatment effects.

The Null Hypotheses

Experimental procedure requires that the effects of an experiment be tested in the form of null hypotheses. That is, that differences occurring in the criterion measure among the treatment groups are not greater than what would be expected by chance. The design of this study is such that a single main effect is to be tested. The null hypothesis for the main effect if: The population mean is the same for all treatments. Stated in terms of the experiment: The mean scores on the objective criterion test are the same for each of three treatment populations. The simple effect null hypothesis is: The differences among corresponding treatment population means are the same for all levels. Again, stated in the terms of the experiment: The differences among the mean criterion test scores are the same between Hi and Mid, Hi and Lo, and Mid and Lo for each treatment. This second null hypothesis concerns itself with the interaction effect of the entire experiment. These null hypotheses apply to the entire population and to each other separately.

In the event that a significant main effect or interaction is found, that is if the first or second or both null hypotheses are rejected, a multitude of further simple effect null hypotheses are possible in order to locate the specific effects. In the event that neither null hypothesis is rejected, further investigation is not warranted.

The Statistical Treatment

The analysis of variance and the "F" test were the main statistical treatments applied to the study. The analysis of variance technique allows the experimenter to analyze in a single double-entry table the effects of various treatments at various levels. The "F" test may then be applied to determine if differences are significant. The assumptions underlying this procedure are:

1. Each treatment group was drawn at random from the corresponding level of the population, and the number drawn was proportional to the number of individuals at that level of the population. This assumption was accepted on the basis of the procedures described in Chapter IV.

2. The distribution of criterion measures for the treatment sub-groups was normal. On the basis of

Norton's[5] study of the "Effects of Non-normality and Heterogeneity on the F-distribution," it was determined that a test of the normality of the distributions was not required. Norton showed that "the F-distribution is practically unaffected by lack of symmetry, per se, in the distributions of criterion measures... In general, the F-distribution seems so insensitive to the form of the distribution of the criterion measure, that it hardly seems worthwhile to apply any statistical test to the data to detect non-normality."[6]

3. Each of the treatment sub-groups had the same variance on the criterion measure. This assumption was tested by the Bartlett test and the results are presented in Chapter IV.

The level of significance for the null hypotheses was set at the 5% level prior to the beginning of the experiment.

As a result of Norton's[7] study showing the need for extreme departure from assumptions 2 and 3, the level of

---

5. Dee W. Norton, "An Empirical Investigation of Some Effects of Non-normality and Heterogeneity on the F-distribution," unpublished Ph.D. dissertation in Education, State University of Iowa, 1952.

6. Lindquist, op. cit, pp. 85-86.

7. Norton, op. cit.

significance for the Bartlett test was set at the 1% level.

Reliabilities and standard deviations were calculated for the criterion measure and the Attitude Inventory in order to make some judgments as to their effectiveness in the study.

The Experimental Procedures

Subjects

As in most psychological experiments, practical considerations played an important role in the selection of the subjects for the study. Two sections of a beginning course in educational psychology were large enough to provide a reasonable number of subjects. The fact that the course was the first education course that the students were taking and that it was a six-credit course (double the weight of most courses) was considered evidence that any grades received would be considered important by the students. The class meetings of both sections were conducted by the same instructors and met on successive hours of the same days: the first at nine A. M. and the second at ten A. M. on Monday, Wednesday, and Friday. The class meetings consisted of a one hour lecture session in an auditorium and a one hour discussion section in small rooms (see course outline Appendix C). The lectures were shared by two regular faculty members and the discussion sections were conducted by graduate student assistant instructors. The total number of students available in the two sections was about 370. However, because of lack of data, failure of individual students to attend key class meetings and late registration for

the course, the actual number of students involved in the study was reduced to 225. There were 99 in the first section and 126 in the second.

The Control Variables

On the first day of the class meeting, the students were asked to answer the items on an Attitude Scale (Appendix A) designed to obtain a measure of their feelings toward persistence in their work.

Following the first class meeting, a list of all students attending was prepared and the grade-point-average for each was obtained from the Registrar of the university. Because GPA's were not available for transfer students, they were dropped from the experiment. The two sections were separately divided into three levels equal in number according to grade point average.

The Treatment Variable

At the beginning of the third week of the term (Monday, the seventh class meeting) the students were told that there would be an essay quiz at the next discussion session meeting on Wednesday. On Wednesday, the students spent the discussion period working on a three-item essay quiz (Appendix A). The items on the quiz were deliberately general so that the students would have difficulty comparing answers after the papers were returned. The students

were told that their quiz papers would be graded and returned at the next discussion period on Friday.

Grades of A, C or F were assigned at random within the levels determined by grade point average. A table of random numbers was used to assign the grades. The digits were assigned grade values as follows:

1, 2 and 3.....A

4, 5 and 6.....C

7, 8 and 9.....F

0.....repeat

A table was entered at random and followed down the columns assigning each paper a grade according to the last digit in each number appearing in the table. For ease of analysis, it was decided to maintain equal Ns in each treatment group. Therefore, once one third of a level had been assigned a particular grade, the corresponding digits to that grade were read as repeats. This procedure was followed until all students within a level had been assigned grades. Those students for whom there was no grade-point-average, were assigned grades of B and D equally so that it would appear to the students that the full range of grades had been used. Ranges and standard deviations for the grade-point-averages appear in Appendix C. This random

assignment of the treatment variable is the basis for the acceptance of the first assumption underlying the statistical treatment of the data (Chapter III). The grades were marked with red pencil on the essay quiz papers. No other marks were made on the papers.

On Friday, the ninth class meeting, the graded papers were returned to the students at the end of the discussion period. The assistant instructors in charge of the discussion periods made no comments about the grades. If pressed by a student, they said that they had not graded the paper and, therefore, could not comment on the grade assigned.

The Criterion Variable

At the beginning of the term, the students had been given an assignment sheet which covered the entire term. A chapter in the text book for the course was the assignment for the tenth class meeting (the Monday following the receipt of the graded essay quiz). An eighteen item objective test (Appendix A) was prepared, covering the assigned chapter. Naturally, this test had not been previously announced. This objective test was administered to the students during the lecture session of Monday's class meeting. The scores on the objective test were intended to be used as the criterion of the effect of the treatment.

The Removal of the Effect

After the test papers were collected, the entire project was explained to the students with an emphasis being placed on the possible effect that grading practices could have on the performance of students. The students were assured that their scores on neither the essay test nor the objective test would be counted in the evaluation at the end of the term.

Summary

A three-by-three treatment-by-levels design was used in the present experiment. The variable under investigation was the arbitrary assignment of grades to students' essays. The control variable which constituted the levels in the experiment was the grade-point-averages of the students. The criterion variable was the scores obtained on an objective test administered after the treatments. Types of error possible in the design were discussed. Null hypotheses were derived for the main effect of the treatments and for the interaction effects. A level of significance of 5% was stipulated for these hypotheses. The statistical technique employed was analysis of variance. The assumptions for this technique were discussed and a level of significance of 1% was established for the test of homogeneity of variance.

The experimental procedures consisted of the application of the treatment of random grade assignment within grade-point-average levels. An objective achievement test was used to measure the effects of the treatment. The experiment was explained to the students after they had completed the achievement test.

CHAPTER IV

INSTRUMENTATION

Two measurement devices were constructed for this study. The first was the Work Persistence Attitude Scale and the second was the multiple choice achievement test used as the criterion instrument. Copies of the two instruments appear in Appendix A. The purpose of Chapter IV is to present data concerning these two instruments.

Work Persistence Attitude Scale

The Work Persistence Attitude Scale consisted of ten items on each of which the students were to rate themselves on a five point scale. Because the lowest score on any item was 1, the possible scale for the instrument extends from 10 to 50. The items were intended to obtain a measure of a student's attitude about work persistence. In order to avoid the intent of the scale being obvious, the items were camouflaged with ten items taken from a standard personality test. The asterisked items on the scale were those scored.

In Table 1, the cell means for the first section are shown. The subjects in each cell were divided into two groups on the basis of criterion test scores. Those students scoring above

TABLE 1

Section 1

Work Persistence Scale Cell Means

Grade Point Average		Arbitrary Grade Assignment A	C	F
Hi	Mean Score for cell	39.2	38.7	35.9
	Mean Score for top 50% on criterion test	39.8	39.8	35.7
	Mean Score for low 50% on criterion test	38.7	37.8	36.2
	Number of cases in cell	11	11	11
Mid	Mean Score for cell	40.5	37.7	37.4
	Mean Score for top 50% on criterion test	40.3	38.0	37.4
	Mean Score for low 50% on criterion test	40.8	37.3	37.3
	Number of cases in cell	11	11	11
Lo	Mean Score for cell	38.8	37.1	38.9
	Mean Score for top 50% on criterion test	39.1	36.9	37.7
	Mean Score for low 50% on criterion test	38.3	37.5	41.0
	Number of cases in cell	11	11	11

Total N	99
Total Mean	38.3
Standard Deviation	3.4
Mean for all cells for top 50% on the criterion test	38.3
Mean for all cells for low 50% on the criterion test	38.3

the mean for the cell were designated the top 50% and those below the mean the low 50%. An inspection of the table shows that the differences between the top 50% and the low 50% are small and not in a consistent direction. As a matter of fact, when the total mean for the table, the total mean for the top 50% and the total mean for the low 50% were compared, they were found to be equal.

The standard deviation of 3.4 on a possible 40 point scale indicates that either the students formed a remarkably homogeneous group or that the measuring device did not function.

To determine which of the two above alternatives was the case, an item count and a test of internal consistency were done using the data from the first section. The counts are presented in Table 2 and the summary of the analysis of variance in Table 3. It can be seen from Table 2 that there was a heavy loading on the high scores. Response scores of 4 and 5 account for more than twice as many responses as 1, 2 and 3 combined.

The estimate of internal consistency of .32 shown in Table 3 is sufficiently low to conclude that the attitude scale is not sufficiently reliable to provide a usable measure.

Further examination of the item counts in Table 3 showed that items 1, 3, 4, 5 and 10 were so heavily weighted toward 4 and 5 responses that they served no purpose in

TABLE 2

Attitude Scale Item Count

Section 1

Frequency of Response Scale Scores

Item Number	1	2	3	4	5
1	0	0	0	21	74
2	1	9	21	42	22
3	0	2	7	34	52
4	1	0	6	31	57
5	0	6	15	42	32
6	5	29	33	19	9
7	27	38	21	9	0
8	3	10	23	39	20
9	4	14	21	35	21
10	0	2	8	38	47
Total	41	110	155	310	334

TABLE 3

Attitude Scale Analysis

Summary Table

Source	ss	df	MS
Total	1293	949	1.36
Individuals	113	94	1.20
Items	491	9	54.56
Error	689	846	.81

$$r = 1 - \frac{\text{MS Error}}{\text{MS Individuals}} = .32$$

TABLE 4

Attitude Scale Analysis for Items 2, 6, 7, 8 and 9

Summary Table

Source	SS	df	MS
Total	674	474	1.42
Individuals	107	94	1.14
Items	107	4	26.75
Error	460	376	1.22

$$r = 1 - \frac{\text{MS Error}}{\text{MS Individuals}} = -.07$$

discriminating among the students. A second analysis was done with these non-discriminating items discarded. Table 4 contains the summary data from the second analysis. The resulting estimate of internal consistency was so close to zero that further efforts to use the results of the Attitude Scale in the analysis of the experiment were considered pointless.

Achievement Criterion Test

The Criterion Test (Appendix A) consisted of 18 objective five choice items. The items covered Chapter 13 of Educational Psychology by Cronbach[1], the text for the course. The statistical characteristics of the test were as follows:

Possible Range	0 to 18
Obtained Range	1 to 17
Variance	16.4
Standard Deviation	4.05
Mean	8.41
Number of chance scores	17
Number of subjects	225

The Kuder-Richardson Formula 21[2] was used to estimate the inter-item consistency of the test with a resulting reliability of .81.

---

1. Lee J. Cronbach, "The Rate of Learning and Forgetting," Educational Psychology, Harcourt, Brace and Company, New York, pp. 379-405.

2. Anastasi, op. cit., p. 110.

The mean score of 8.41 with a standard deviation of 4.05 shows an average difficulty level for the test of slightly above 50%. The obtained range of 17, the distribution of scores seen in Table 5 and the low number of chance scores (of which only seven were found in the Lo G. P. A. level) indicate that there was neither a ceiling nor a floor effect resulting from the test itself. Finally, the reliability estimate of .81 is an acceptable level for an 18 item test. The Criterion Test mean scores for the cells shown graphically on page 26 appear in Appendix B.

The Bartlett[3] test of homogeneity of variance was applied to the Criterion Test data in order to make a judgment about the third assumption for analysis of variance expressed in Chapter III. The level of significance required to reject the null hypothesis that the variances are equal was set at 1%.

The Bartlett[4] test employs the X^2 distribution to determine significant differences of variance. The 1% value of X^2 for a-1 degrees of freedom (9-1=8) is 20.090. The obtained value for X^2 from the data was $\frac{16.825}{C}$ (C being a constant greater than 1). Hence, the null hypothesis concerning

---

3. H. O. Hartley, "Testing the Homogeneity of a Set of Variances," Biometrika, V. 31, (1940), pp. 249-255.

4. Ibid.

TABLE 5

Distribution of Achievement Criterion Test Scores

Frequency

Scores	5 10 15 20 25 30	Number
18		0
17	.	1
16	....	4
15	.....	8
14	.....	9
13	.....	12
12	.....	10
11	.....	19
10	.....	18
9	.....	25
8	.....	29
7	.....	16
6	.....	24
5	.....	14
4	.....	19
3	.....	10
2	.....	5
1	..	2
0		0

homogeneity of variance may not be rejected and the third assumption concerning the analysis of variance may be accepted.

On the basis of the above characteristics, it was concluded that the Criterion Test was capable of reflecting the effects of the treatments applied prior to its administration.

Summary

A careful analysis of the Work Persistence Attitude Scale revealed an extremely limited distribution with a low reliability. Manipulations of the scale to improve its usefulness were unsuccessful. It was determined that the scale did not function, and it was therefore not used in further analysis.

The Criterion Test was examined and found to have an acceptable distribution and a reliability of .81. These, along with other characteristics, indicated that it was capable of reflecting treatment effects. The Bartlett test for homogeneity of variance was applied to the Criterion Test scores. It resulted in the acceptance of the assumption of homogeneity.

CHAPTER V

ANALYSIS OF THE RESULTS

The analysis of the results was done with three different sets of data. The first included all subjects in the experimental group. The second and third analysis were done separately, for each sex.

As explained in Chapter III, the technique employed was that of the analysis of variance in a treatment by levels design. The mean scores for the cells are shown in Appendix B. The actual analysis was done with the second class section treated as a replication, thus making the analysis a three treatment by three levels with one replication design (3 x 3 x 2). The data resulting from the analysis are presented in Table 6.

The main effect null hypothesis is concerned only with the three treatment group populations and, therefore, the test using the F distribution was performed by finding the ratio of the mean square for treatments to the mean square within cells.

$$F = ms_T \; / \; ms_W = 1.05$$

5% value of F for 2 and 223 df = 3.04

Therefore, the main effect null hypothesis may not be rejected.

Sour	
Level	
Treat	
Secti	
L x T	
S x L	
T x S	
L x	
Within	
Total	

TABLE 6

Summary Table

Analysis of Variance of Scores on the Achievement Criterion Test

All Subjects

Source	df	ss	ms
Level (L)	2	309	154.5
Treatment (T)	2	23	11.5
Section (S)	1	12	12.0
L x T	4	20	5.0
S x L	2	5	2.5
T x S	2	13	6.5
L x S x T	4	6	1.5
Within cells (w)	223	2426	10.9
Total	240	2814	

The simple effect null hypothesis is concerned with the treatment having different effects at different levels. This is the interaction effect and is tested with the ratio of the mean square for L x S x T and the within cells mean square.

$$F = ms_{LST} / ms_{w} = 0.14$$

5% value of F for 4 and 223 df = 2.41

Thus, the simple effect null hypothesis was not rejected, and, as stated in Chapter III, further investigation of the data for significant effects was not warranted.

The same techniques were used in the separate analyses by sexes. In Table 7 the data is presented for females followed by the tests of the null hypotheses. Table 8 contains the data for males followed by the appropriate tests.

In neither case, males nor females, do the F ratios warrant rejection of the null hypotheses.

The replication data were included in the analysis of variance summarized in Table 6. The F Test was used to establish that the variance of the second section did not differ significantly from that of the first section. Stated as a null hypothesis: The population mean is the same for both sections. The ratio for this test is that of the mean square for section (S) and the mean square within cells (W).

$$F = ms_{S} / ms_{W} = 1.10$$

5% value of F for 1 and 223 df = 3.89

TABLE 7

Summary Table

Analysis of Variance of Female Scores

on the Achievement Criterion Test

Source	df	ss	ms
Level (L)	2	145	72.50
Treatment (T)	2	2	1.00
(Cells)	(8)	(175)	
L x T	4	28	7.00
Within Cells (W)	50	691	13.82
Total	58	866	

Tests for significance:

Main Effect $F = ms_T / ms_W = .07$

5% value of F for 2 and 50 df = 3.18

Simple effect $F = ms_{LxT} / ms_W = .51$

5% value of F for 4 and 50 df = 2.56

Thus, the null hypothesis was not rejected and the replication was considered to have substantiated the results found with the first section.

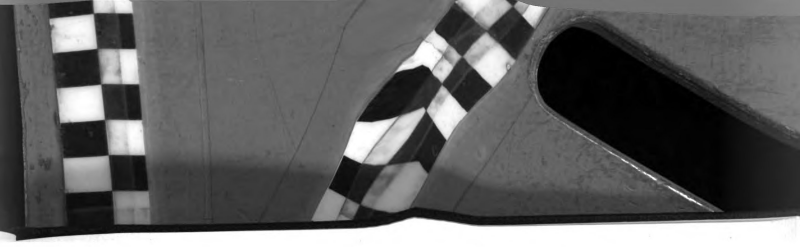
Cell means for the replication are shown in Appendix B.

Summary

The analysis of variance was performed as a 3 x 3 x 2 including the replication with the second section of the class. The F tests of the null hypotheses were not significant at the stated level of confidence, and therefore the null hypotheses were not rejected.

The separate analyses done for each sex failed to produce a significant F ratio for either of the null hypotheses.

The F test comparing the variances between the two sections showed no significant difference, and it was therefore accepted that the replication section substantiated the results.



CHAPTER VI

SUMMARY AND CONCLUSIONS

Summary

The Problem

The purpose of the study was to investigate the effects of the arbitrary assignment of high or low grades on students' performance on their following assignment. A theoretical structure was developed based on positive and negative reinforcement. It was hypothesized that performance subsequent to the treatment would be influenced by the reinforcement, by the intensity of the reinforcement and by students' attitudes toward their work.

A review of the pertinent literature revealed that not only was little known about the effects of grades, but that little effort had been made to investigate the subject. The one study which had been done with reasonable controls indicated that a single assigned grade was capable of influencing attitudes toward a specified subject.

The Design

The design uses the three elements of modern experimental procedure: randomization, control and replication. The

treatment-by-levels design was determined to be the most appropriate for the study because of the likelihood that the criterion measure would be correlated with students' academic ability. There were three treatments and three levels, forming a 3 x 3 design. The levels were formed by dividing the subjects into thirds based on their grade point averages. The treatments consisted of the random assignment within levels of A, C or F grades to essay tests. A replication was included and provision was made for separate analyses by sex. The analysis of variance statistic was employed in the analysis of the data. A confidence level of 1% was set for testing assumptions and 5% for testing the results of the analysis of criterion test scores.

Instruments

The Work Persistence Attitude Scale was designed for use as a control instrument. The items were intended to elicit responses indicating the students' attitudes toward persisting in their work when confronted with frustration, interruption or failure. The reliability of this scale was estimated to be .32, which was considered to be a level such that use as a control variable in a separate analysis was unwarranted.

An 18-item multiple-choice achievement test was constructed covering a single chapter of the course text book.

It was designed for use as the criterion instrument intended to reflect the effects of the treatments. The reliability was estimated to be .81, the mean was 8.41 and the standard deviation was 4.05. These characteristics were interpreted as indicating that the test was capable of reflecting treatment effects.

Experimental Procedures

The experiment was conducted in ten sections of a beginning Educational Psychology course at Michigan State University in the Fall of 1959. The total population of the experimental group, including the replication, was 225.

The Work Persistence Attitude Scale was administered at the first class meeting of the course. Grade point averages were collected from the university registrar. The levels by grade point average were established, a treatment of randomly assigned grades on an essay test was applied during the third week of the course. The Criterion Test, covering the text book assignment was given three days after the treatment. Immediately following the administration of the Criterion Test the experiment was explained to the students as a learning experience in educational psychology.

Tests of the Null Hypotheses

The tests of the hypotheses using the previously established 5% level of confidence for the F ratio were all negative. Main effects (treatments alone) were tested for the entire group and for each sex separately. In no case was the null hypothesis rejected. Simple effects (interaction of treatments by levels) were similarly tested and again the null hypotheses could not be rejected.

The null hypothesis stating that there was no difference between replications was tested, and again the resulting F ratio was not sufficiently high for rejection.

Conclusion

The failure to reject any of the null hypotheses was interpreted as indicating that within the limitations of this study the random assignment of grades within grade point levels had no effect on the students' performances on their following assignment. Generalization from these results is curtailed because of the limitations discussed below and because the failure to reject a null hypothesis does not automatically imply acceptance. It is possible, however, to use the results and findings to

(1) consider other research which should be done to more fully explore the problem and

(2) temper a hasty conclusion that either grading hard or easy makes much difference.

Limitations of the Study

1. The first limitation involves the treatments. Repeated application of the treatments in order to acquire greater reinforcement was determined to be unethical because the experiment was being conducted in a real situation without volunteers. The problems involved in removal of possible experimental effects might have proved unmanageable. For example, an average student repeatedly receiving F's in a six-credit course in his major might leave the university before effort to remove the experimental effects had been made.

2. The second limitation was the age and experience level of the subjects involved. Students who have reached the sophomore level in college have had much experience in surviving in educational situations. Thus, the advanced academic level in which the experiment was conducted was a severe limitation when coupled with the ethical decision, mentioned previously, not to repeat the treatment and thus gain greater reinforcement.

3. The most serious limitation within the study itself was the failure of the Work Persistence Attitude Scale to provide meaningful material which could be used to discriminate among students about their relative susceptability to positive and negative grade reinforcement.

Another limitation was the dropouts from the starting population. The reasons for the dropouts could have been significant. For example, it was necessary that a student attend class

on all three days of the experiment; write the essay quiz on the Wednesday, receive the graded essay quiz on Friday, and appear on Monday to take the objective quiz. It was hypothesized that the grade received would influence the students' academic behavior. Ifthe influence resulted in not doing the weekend assignment and, as a result, a decision not to attend class on Monday, it is possible that a part of the population was lost that, if included, would have caused a significant result.

A Fifth limitation was the possibility that the treatment was not internalized by enough students to make any difference. That is, the students may simply not have believed that the grades on the essay quiz were important and thus, did not behave differentially according to the grades received.

The final limitation was also a matter of the subjects. Any generalization resulting from the study would have to be limited by the fact that all subjects were voluntarily enrolled in a course leading to a teaching certificate. What biases this fact may have lent to the study are unknown.

Implications for Further Research

The implications for further research are actually a reiteration of the limitations of the present study. That further research is needed has been testified to in the review of the literature where the reviewers of the subject in the past

two editions of the Encyclopedia of Educational Research[1] stated that little had been done and much more is needed. The implications are for both a refinement and expansion of this study:

1. The development of a sensitive and sophisticated instrument for the measurement of student attitudes toward the grades they receive is needed for use as a control in future studies attempting to measure effects of grades on student performance.
2. Studies are needed at various age and experience levels in order to determine where the grades assigned to students' work are most influential. As discussed in the limitations, it may be that students at the college level have received so much reinforcement toward a particular grade level that no short term treatment can be effective in influencing academic performance.
3. Although no specific recommendation is offered, a need exists for an ethical approach to experimentation in which the treatment can be applied repeatedly for maximum effect.
4. There is a need for studies of the effects of grades on other than academic achievement.

---

1. Odell, op. cit., pp. 711-717.
   Wingo, op. cit., pp. 848-859.

Efforts should be made to answer questions about the effects the grades assigned to students have on

(a) their attitudes toward school,

(b) their feelings toward other students,

(c) their attitudes toward any academic pursuit and toward specific curriculum areas,

(d) their approaches to other kinds of work, and

(e) their self concepts in general.

BIBLIOGRAPHY

Anastasi, Anne, Psychological Testing, MacMillan, New York, p. 110, "Test Reliability".

Barton, W. A., Jr., "Pupil Reaction to School Reports, " The School Review, 33:771-780; 1925, 34:42-53, 1926.

Book, W. F. and Norvell, L., "The Will to Learn", Pedagogical Seminary and Journal of Genetic Psychology, 29: 305-362; 1922.

Bostrom, R. N., "Grades as Reinforcing Contingencies and Attitude Change", Journal of Educational Psychology, April, 1961, V. 52, No. 2.

Cartwright, Dorwin, "The Effect of Interruption, Completion, and Failure Upon the Attractiveness of Activities", Journal of Experimental Psychology, 15:73-86; 1932.

Cronbach, Lee J., "The Rate of Learning and Forgetting," Educational Psychology, Harcourt, Brace and Company, New York, pp. 379-405.

Doob, L. W., "The Behavior of Attitudes," Psychological Review, 1947, 54:135-156.

Edmiston, R. W., and Jackson, L. A., "The Relationship of Persistance to Achievement", Journal of Educational Psychology, 1949, 40:47-51.

Edwards, A. L., Experimental Design in Psychological Research, (Revised Edition) New York: Rhinehart, 1960.

Encyclopedia of Educational Research, New York, McMillan, 1960.

Encyclopedia of Educational Research, New York, McMillan, 1950

Farquhar, William W. and Krumboltz, John D., "A Check List for Evaluating Experimental Research in Psychology and Education," Journal of Educational Research, 52: 353-354, 1959.

Fay, P. J., "The Effect of the Knowledge of Marks on the Subsequent Achievement of College Students, " Journal of Educational Psychology, 28: 548-554; 1937.

Fischer, R. P., "The Role of Frustration in Academic Under-Achievement: An Experimental Investigation," Journal of the American Association of College Registrars, 1943, 18: 227-238.

Gebhard, Mildred E., "The Effect of Success and Failure Upon the Attractiveness of Activities as a Function of Experience, Expectation and Need," Journal of Experimental Psychology, 38: 371-388; 1948.

Hartley, H. O., "Testing the Homogeneity of a Set of Variances," Biometrika, V. 31, (1940), pp. 249-255.

Hilgard, Ernest R., "Reinforcement, Provisional Try, and Feedback," Theories of Learning, Appleton-Century-Crofts, Inc., New York, pp. 469-472.

Horrall, B. M., "Relationships between College Aptitude and Discouragement--Buoyancy Among College Freshmen," Journal of Genetic Psychology, 1949, 72: 185-243.

Hurlock, E. B., "An Evaluation of Certain Incentives Used in School Work," Journal of Educational Psychology, 16: 145-159; 1925.

Lindquist, E. F., "Treatments X Levels Designs," Design and Analysis of Experiments in Psychology and Education, Houghton Mifflin, Boston, pp. 121-149.

Little, Ruth C., "Whither Grading," National Educational Association Journal, 36: 12-13; 1947.

Norton, Dee W., "An Empirical Investigation of Some Effects of Non-normality and Heterogeneity on the F-distribution," unpublished Ph.D. dissertation in Education, State University of Iowa, 1952.

Odell, C. W., "Marks and Marking Systems," Encyclopedia of Educational Research, New York, McMillan, 50, pp. 711-717.

Page, Ellis Batten, "Teacher Comments and Student Performance: A seventy-Four Classroom Experiment in School Motivation," Journal of Educational Psychology, August, 1958, V. 49, No. 4.

Portwood, T. B., "Success does it--When Pupils Can't Fail, They Don't," Nations Schools, 28: 60-62; 1941.

Sears, Pauline S., "Levels of Aspiration in Academically Successful and Unsuccessful Children," Journal of Abnormal and Social Psychology, 35: 498-536; 40.

Sears, Robert R., "Success and Failure," in McNemer, Quin, and Merrill, Maud A. (Eds.), Studies in Personality, McGraw Hill, 42: pp. 235-238.

Taylor, H. R., and Constance, C. L., "How Reliable are College Marks?," Research and Higher Education, U. S. Office of Education Bulletin, 1931, No. 12, pp. 5-14.

Tiege, E. W., Tests and Measurements for Teachers, Houghton, 1931, pp. 170-171.

Wingo, G. Max, "Methods of Teaching," Encyclopedia of Educational Research, New York, McMillan, 60, pp. 848-859.

APPENDIX A

INSTRUMENTS

Essay Quiz

Work Persistence Attitude Scale

Criterion Test

Essay Quiz

FE 200

1. Briefly explain the contribution of psychology to the field of education.

2. Present a plan for teaching a particular concept (choose your own, be specific as to what the concept is).

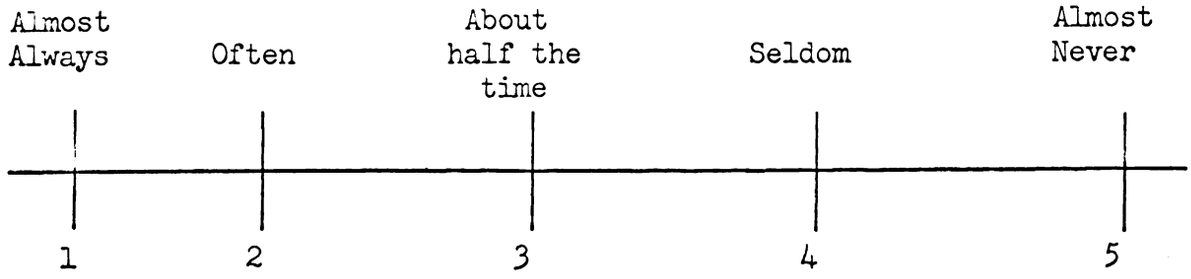
3. Discuss the teaching of attitudes (or striving for attitude changes) in the school.

ATTITUDE SCHEDULE

Directions

This schedule consists of a number of statements to which you are asked to react by rating your feelings toward them on a five point scale.

You are to record your responses to the statements on the answer sheet provided. Use the following scale to indicate your feelings about the statements.



You will notice that the numbers below the descriptive phrases (Almost always, Often, Seldom, etc.) correspond to the numbers on your answer sheet. Thus, if you wish to respond to an Item by saying "Almost Never" blacken the space on your answer sheet below the number 5.

Example:

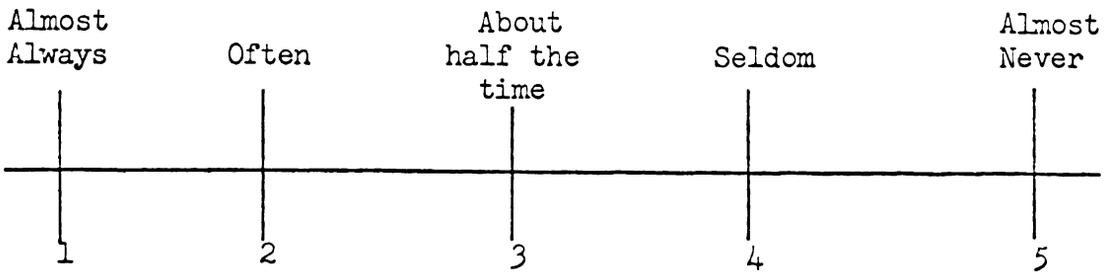
1. I like to eat steak.

If you almost never like to eat steak, you should mark your answer sheet as follows:

Be careful to keep in mind that each statement must be evaluated according to its own meaning. Thus, in the example, you might answer "Almost Always", even though you probably wouldn't want to eat steak twenty-four hours a day. Rather, your answer of "Almost Always" would indicate that at appropriate meals (most likely supper), you would prefer steak for the main course.

The same principle applies to each of the statements on the schedule.

Name_____



* 1. I like to finish any task that I begin.

2. I feel like telling people off when I disagree with them.

3. I like to experience novelty and change in my daily work.

* 4. When I cannot seem to get anywhere with a tough job or problem, I like to forget about it until I have to come back to it again.

5. I like to conform to custom and to avoid doing things that people I respect might consider unconventional.

* 6. When I receive a low grade early in the term, I am naturally discouraged and feel like not trying anymore.

7. I like to forgive my friends who may sometimes hurt me.

* 8. When I have an assignment to do, I like to start in and keep working on it until it is completed.

9. I like to be independent of others in diciding what I want to do.

*10. If I receive a high grade early in the term, I naturally feel complacent and have a tendency to ease off in my work.

11. I like to rely on my ability to react to the immediate situation in order to accomplish my aims.

*12. I resent being interrupted while at my work.

13. I feel better when I give in and avoid a fight, than I would if I tried to have my own way.

*14. When I fail to accomplish what I want, I try to find something else with which I can be successful.

15. I feel that I should confess the things that I have done that I regard as wrong.

*16. An experience offailure makes me hesitate to try the same thing again.

17. I like to think about the personalities of my friends and to try to figure out what makes them as they are.

*18. I like to stay up late working in order to get a job done.

19. I like to go out with attractive persons of the opposite sex.

*20. When I am successful, I try even harder the next time.

FE 200

Objective Quiz

1. Pleasant associations are retained (learned) longer than unpleasant ones. This is explained by:

   * a. It hasn't been shown how this aids learning.
   b. Unpleasant things are associated with discomfort and and are deliberately forgotten.
   c. The phenomena of retroactive inhibition.
   d. Pleasant associations have an inherent quality which makes them easier to retain.
   e. The anti-retention quality of unpleasant associations.

2. Spacing of review sessions should generally be

   * a. distributed with an increasing time between sessions.
   b. distributed with a decreasing time between sessions.
   c. concentrated early in the learning situation.
   d. concentrated late in the learning situation
   e. distributed evenly throughout the learning situation.

3. A leveling off of a learning curve should be interpreted by the teacher as

   a. a lagging in motivation with the need for rewards.
   b. too much concentration on speed.
   c. a limit of the students ability.
   d. over concentration on immediate goals.
   * e. a plateau with diagnosis required.

4. The period of "decreasing gains" in a learning curve is indicative of

   a. lack of concentration by the subject.
   b. an approach to the subjects capacity.
   c. the presence of distractions in the learning situation.
   * d. refinement of gross improvements.
   e. lack of understanding of the task.

5. In the study which tested how much of a psychology lecture was remembered the period of greatest forgetting was

   * a. during the first week.
   b. between the first and second week.
   c. divided equally during the first three weeks.
   d. after the end of the second week.
   e. after the end of the first week.

6. Performance records after limited practice compared to final standings show

   a. no correlation from one to the other.
   * b. that a person tends to have somewhat the same rank in his group.
   c. a high negative correlation.
   d. that early records can be used for individual prediction purposes.
   e. the importance of motivating techniques.

7. The period of "negligible progress" indicates that

   a. the individual has reached the maximum of consumption.
   b. little or no learning is taking place.
   * c. readiness skills are developing.
   d. retroactive inhibition is occuring.
   e. the individual is resting.

8. A likely explanation for the results of the retention experiment with pro-Communists and anti-Communists is

   a. people remember things which are completely contrary to their own beliefs.
   b. Anything that is painful is remembered vividly.
   * c. People recall ideas that fit their basic ideas of what is true and reasonable.
   d. First impressions are most significant in determining retained material.
   e. People remember things which are shocking to them.

9. Overlearning is a type of learning that the teacher

   a. should discourage.
   * b. should strive for in selected areas.
   c. should work for with special students.
   d. should limit to motor skills.
   e. should limit to language skills.

10. Studies of retention show that

    a. an extreme amount of forgetting is inevitable.
    * b. it is possible to actually have gains after the end of instruction.
    c. a fair amount of forgetting is bound to occur and the things forgotten must be relearned.
    d. the amount of interference can be used to predict the amount of forgetting.
    e. overlearned material is forgotten as quickly as any other.

11. In Bartlett's study on recall the sketch which was reproduced from memory

    a. left out many details.
    b. was more irregular than the original.
    * c. was more regular than the original.
    d. showed considerable interference.
    e. lacked smoothness.

12. Plateaus result from a number of causes. Which of the following is NOT one of these causes?

    a. The subject is making small changes which do not show in the measurement.
    b. The subject is no longer changing his method of performance.
    c. The subject is not changing in his physical development rapidly enough to show.
    d. The subject is lacking in sufficient comprehension of the task.
    * e. The subject has forgotten some of the basic rudiments of the task.

13. In Moser's study which involved teaching fractions to second graders, a test given several weeks after the end of instruction showed

    a. that the children had almost entirely forgotten the subject.
    b. a fifty per cent drop in performance.
    c. no significant difference in performance.
    * d. an improvement over previous performance.
    e. that other learning had interfered.

14. An over-all learning curve omits

    a. abrupt changes in speed of improvement.
    * b. diagnostic information.
    c. the early period of little improvement.
    d. summarized information.
    e. the approach to limit.

15. To avoid interference a teacher should

    a. avoid difficult material in the beginning.
    b. teach different reactions to situations that seem similar to the learner.
    c. cover as broad an area as possible.
    d. challenge students with provacative material.
    * e. provide reinforcement of generalizations until they are mastered.

16. According to Cronbach, very often forgetting occurs because

    a. memories fade from the mind much as colors fade in a fabric.
    b. the mind can only retain a certain amount of material.
    c. the individual has not received enoughmemory training.
    * d. one set of ideas has been mixed up with another causing interference.
    e. the individual was not properly motivated in the original learning situation.

17. A person's poorer performance after some period of time without practice is <u>not</u> caused by which of the following:

    a. Errors of reorganization.
    b. Interference from conflicting learning.
    * c. Errors of over-interpretation.
    d. Errors of over-simplification.
    e. Unorganized or incorrectly organized original learning.

18. Retention in a high school chemistry course was highest in

    a. technical vocabulary.
    b. balancing equations.
    c. specific facts about compounds.
    * d. application of principles.
    e. formula construction.

APPENDIX B

Criterion Test Cell Means

CRITERION TEST

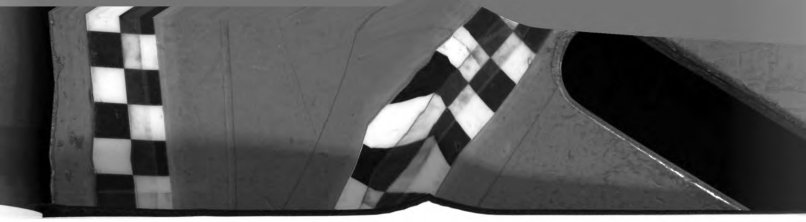
CELL MEANS

First Hour Group

Treatment

Essay Test Assigned Grade

Level Grade Point Average		A	C	F
	Hi	10.64	10.09	10.90
	Mid	8.45	7.27	8.09
	Lo	8.36	6.64	7.64



CRITERION TEST

CELL MEANS

Second Hour Group

Treatment

Essay Test Assigned Grade

Level Grade Point Average	A	C	F
Hi	8.93	9.36	10.71
Mid	8.00	7.57	7.43
Lo	7.21	6.93	7.71

APPENDIX C

Course Outline

FE 200 Section 2

The Individual and the School

Spring 1959

I. Purpose

FE 200 is the first in a series of planned experiences designed to help you become an effective teacher. The course is essentially an introduction to educational psychology and, as such, emphasizes the understanding of psychological principles as they apply to education. More specifically, the course should help you

A. learn the basic principles of how children grow and develop and how environment and maturation affect their behavior.
B. learn general methods of assisting children to learn optimally.
C. learn the psychological factors that promote and inhibit efficient learning.
D. learn the important principles of measuring readiness for learning and the outcomes of learning.
E. become familiar with the resources and methods of solving educational problems.
F. appreciate and understand how the scientific method provides new knowledge of human behavior.
G. develop attitudes toward children and toward the profession of teaching that will enable you to be maximally effective.

II. Appointments with Instructors

Each instructor will be available for individual consultation during the office hours below -- other times by appointment only. Offices are located on the fourth floor of the College of Education Building. The receptionist on duty will assist you by finding out if the instructor is available and directing you to his office. Call EXT. 3561 for appointments.

Instructor	Days	Hours
Mr. Lewis	MWF	8:30 - 9:30
Mr. Nolan	T TH	9:30 -11:00
Mr. Vander Linde	MTWTh	8:30 - 9:30
Dr. Farquhar	MT	8:30 - 9:30
Dr. Krumboltz	MWF	8:00 - 9:00

III. Readings

Textbook for the course is Cronbach, L. J. Educational Psychology, N. Y.: Harcourt Brace, 1954.

Assignments will also be made in the following references available in the Assigned Reading Room of the Library.

Call No.	Author	Title
LB 1051 .54	Seidman	Readings in Educational Psychology
LB 1051 .C343	Coladarci	Educational Psychology, A Book of Readings
152.27 c 287M 1954	Carmichael	Manual of Child Psychology
152.27 D 411r	Dennis	Readings in Child Psychology
LB 1051 .F97	Fullagar	Readings for Educational Psychology
150.8 K96	Kuhlen	Psychological Studies of Human Development
LB 1105 .R4	Remmers	Growth, Teaching, and Learning

IV. Summary of Course Requirements and Evaluation Procedures

The course requirements are designed to help you attain the objectives of the course. Your grade will be a measure of how well you have attained the course objectives. A variety of procedures will be used to assess your progress. Your final grade is a measure of the quality, not the quantity, of your work.

Contribution to Class Grade	Method	Explanation in Section
25%	Midterm examination	X
10%	Short quizzes (unannounced)	X
20%	Written Report	VIII
10%	Report on field trip	IX
10%	Journal Reports	VI
15%	Instructors' subjective estimate of your potential for becoming a good teacher as judged from the quality of your class participation and activities.	VII
10%	Class final examination	X

Your class grade will be determined in the above manner. In addition you will also receive a grade from the FE 200 Common Final Examination. These two grades will then be weighted together (the exact contribution of each has not yet been decided) to form your final grade.

You should plan on 4 hours of examinations during examination week for a 6-credit course. Two hours will be devoted to the common final examination at a time to be announced, probably Friday, June 12, 10:15 - 12:15. The other two hours will be on Friday, June 12, 8:00 - 10:00 A. M.

V. Course Outline

Each assignment is to be completed before class on the date indicated.

Part I

Human Growth and Development

General Objectives

A. To learn the basic principles of how children grow and develop and how maturation affects their behavior.
B. To appreciate and understand how the scientific method provides new knowledge of human behavior.

Date	Topic	Assignment
April 6	Effective group discussion Types of group members Steps in problem solving	Cronbach, Chapter 1 Seidman #14 "Research in Group Behavior"
April 8	Study Techniques: An Overview of Applied Learning How to make notes How to prepare for examinations How to structure the study situation	Coladarci #3 "The Relation of Schools of Psychology to Educational Practice
April 10	How we Study Development Longitudinal Cross-Sectional Experimental Clinical Principles of Development	Cronbach Chap. 2 Fullagar #12 (or Coladarci #6) "Concepts of Growth"
April 13	Overview of Development Pre-school Middle years	Cronbach Chap. 3
April 15	Overview of Development Adolescent Adult	Cronbach Chap. 4 Remmers pp. 107-109 "When are children ready to learn?"
April 17	Individual Differences Readiness Extent of individual differences	Cronbach Chap. 5 First Journal report due

Part II

Learning

General Objectives

A. To learn general methods of assisting children to learn optimally.
B. To become familiar with the resources and methods of solving educational problems.
C. To appreciate and understand how the scientific method provides new knowledge of human behavior.

Date	Topic	Assignment
April 20	Semantics and the Objectives of Education Language may be used for many purposes Review of the ladder of abstraction Formulating operational definitions of educational objectives	Cronbach Chap. 14
April 22	Transfer of training Achieving educational objectives Theory of formal discipline Theory of identical elements Role of scientific method Teaching for transfer	Cronbach Chap. 9 Remmers pp. 170 "What is the Transfer Value of Certain School Experiences?"
April 24	Motivation What makes people tick? Psysiological and psychological needs Intrinsic and extrinsic motivation	Coladarci #23 "Motivation Theory and Educational Practice"
April 27	Learning concepts and principles What is a concept How we form concepts	Cronbach Chap. 10 Coladarci #32 "Rote Memorization, Understanding and Transfer
April 29	Learning concepts and principles How we use concepts How we build correct concepts	Cronbach Chap. 11 Coladarci #26 "Success and failure in the classroom"
May 1	Practice and learning of skills What is skilled action A demonstration Methods of effective skill building	Cronbach Chap. 12 Second Journal Report due
May 4	Teaching for permanent learning Factors related to retention A frame of reference Implications for teachers	Cronbach Chap. 13 Remmers pp. 110 "A Study of Retention of Classroom Learning"
May 6	Midterm Examination	Review

May 8	Discipline: The Beginning Teacher's Bugaboo Conditions for effective discipline Constructive suggestions Preventive steps in traditional school	Cronbach Chap. 15 Fullagar #49 "Teacher behavior liked and disliked by pupils"
May 11	Examination of the examination An opportunity to check your strengths and weaknesses by reviewing your performance on the midterm test.	Fullagar pp. 57-66 (or Remmers pp. 99-106) "Are Theories of Learning Helpful?"

Part III

Personality Development and Character Education

General Objectives

A. To learn the psychological factors that promote and inhibit efficient learning.
B. To appreciate and understand how the scientific method provide new knowledge of human behavior.
C. To develop attitudes toward children and toward the profession of teaching that will enable you to be maximally effective.

Date	Topic	Assignment
May 13	Mental Hygiene Sources of help Psychologist Psychiatrist Social Worker Psychiatric social worker Counseling	Cronbach Chap. 17 Coladarci #45 "Charting Social Relationships of School Children"
May 15	Dynamics of Adjustment Mechanisms The feeling of inadequacy	Cronbach Chap. 18 Seidman #60 "Multiple response to frustration" Third Journal Report due
May 22	Evaluation of Personality Unstandardized devices Autobiography Sociometric techniques Anecdotal records	Field Trip Report due

May 25	Standardized devices Objective tests Projective tests	Cronbach Chap. 19

Part IV

Educational and Psychological Measurement

General Objectives

A. To learn the important principles of measuring readiness for learning and evaluating the outcomes of learning.
B. To become familiar with sources which may help the teacher to solve problems in measurement.

May 27	How you can measure learning objectively Purpose of measurement Writing objective tests	Cron bach Chap. 6 Seidman #27 "The Forms of Objective Test Items"
May 29	Are essay tests more valuable? The argument Constructing essay questions The Problem of scoring	Write and bring to class two superior objective test items. Fourth Journal Report due
June 1	Concepts needed for understanding tests Central tendency Variability Normal distribution Derived scores	Cronbach Chap. 16 Seidman #23 "Suggestions for improving essay questions"
June 3	Is the test any good? Correlation Reliability Validity Objectivity	Cronbach Chap. 7
June 5	Assigning marks Consider your objectives again Issues involved A point of view	Cronbach Chap. 8 Seidman #61 "Reporting pupil progress" Written Report due
June 8	Dean Melby: "The Profession of Teaching"	Coladarci #22 (or Remmers pp. 194-199) "Individual differences and curriculum practices"

June 10	Putting the pieces together Theory and Fact in Educational Decision Making Your Future in Education Mothod of Final Evaluation
June 12	Final Examinations 8:00 - 10:00 10:15 - 12:15

VI. Instructions for Journal Reports

A. Each student will report on four articles from the periodicals listed below. Each article will be from a different periodical.
B. One report will be due on each of the dates indicated in the Course Outline, Section V.
C. Reports must be concerned with some phase of educational psychology. You may have to browse through several articles before you find a good one. Perhaps some of your articles will be directly related to the subject matter you intend to teach.
D. Reports must be typewritten and follow the format shown on the attached form.
E. Articles will be selected which satisfy the following criteria:
   1. Periodicals will be from attached list.
   2. Subject matter may be theoretical or practical, but it must have implications or applications for your future teaching. If it does not, then you must find another article which does have some implications for you.
   3. You must be able to understand the subject matter of the article. Don't try to report on an article that is too advanced for your present training.
F. Student reports will be evaluated on the following basis:
   1. Quality of summarization of the article: conciseness, comprehensiveness, clearness.
   2. Appropriateness of article to your interests and to the field of educational psychology.
   3. Quality of statements of implications and applications: concreteness, realism.
   4. Neatness, following prescribed form: English usage, etc.

G. List of periodicals available in the Library:

Adult Education
Agricultural Education Magazine
American Biology Teacher
American Business Education
American Childhood
American Journal of Physics
American Teacher's Magazine
American Vocational Journal
Arithmetic Teacher
Athletic Journal
Balance Sheet
American Educational World
California Journal of Secondary Education
Camping Magazine
Catholic Education Review
Catholic School Journal
Child Development
Child Study
Childhood Education
Clearing House
Education
Educational Administration and Supervision
Educational and Psychological Measurement
Educational Digest
Educational Forum
Educational Leadership
Educational Outlook
Educational Record
Educational Research Bulletin
Educational Screen and Audio-Visual Guide
Elementary English
Elementary School Journal
Elementary School Teacher
English Journal
Exceptional Children
French Review
German Quarterly
Grade Teacher
High School Journal
Hispania
Industrial Arts and Vocational Education
Journal of Business Education
Journal of Chemical Education
Journal of Education
Journal of Educational Psychology
Journal of Educational Research

Journal of Educational Sociology
Journal of Engineering Education
Journal of Experimental Education
Journal of General Education
Journal of Geography
Journal of Health, Physical Education, and Recreation
Journal of Home Economics
Journal of School Health
Journal of Negro Education
Journal of Speech and Hearing Disorders
Journal of Teacher Education
Mathematics Teacher
Michigan Education Journal
Modern Language Journal
Music Journal
Journal of the National Education Association
National Elementary Principal
National Parent-Teacher
Nations' Schools
Nervous Child
Personnel and Guidance Journal
Practical Home Economics
Progressive EDucation
Quarterly Journal Of Speech
Review of Educational Research
Religious Education
Safety Education
School Activities
School and Society
School Arts
School Executive
School Life
School Musician
School Review
School Science and Mathematics
Scholastic Coach
School Shop
Science Education
Science Teacher
Social Education
Social Studies
Understanding the Child
Journal of Counseling Psychology
The Educational Record

Periodicals not on this list must be approved by one of the Instructors.

JOURNAL ARTICLE REPORT

FE 200 Section 3

Name: John D. Doe Room 109

Report No. 3 Date November 7, 1958

Allen, Edward D., "How to Teach Students to Think in Spanish", The Modern Language Journal, March 1958, Vol. 52, pp. 139-41.

My Summary of the Article:

Purpose: The purpose of this article is to show how a student can be taught to think in Spanish by applying the principles of reflective thinking or the scientific method to their problems.

Method: The first rule of the classroom was to speak no English whatsoever. The teacher either acted out the words or wrote on the board. Concrete vocabulary did not present a problem, but abstract words had to be taught by using sentences whose meanings could be ascertained through inference. For example -- the Spanish word "trabaja" means "works" in English. To put the meaning of this word across to the students, the teacher wrote seven different sentences on the board all of which contained the word "trabaja". e. g.

1. El Senor Allen trabaja en la escuela.
2. El Senor Eisenhower trabaja en Washington.
3. El estudiante que trabaja mucho recibe una "A". ....

From these sentences the students had to solve the problem of what "trabaja" meant. First of all, they set up several hypotheses and proceeded to test each one. They found one which suited all the conditions;and thereby discovered the true meaning and could make the generalization that "trabaja" meant "work". They could now apply this "meaning-carrier" to new situations.

Findings: Although it may have been easier for the teacher to say "trabaja-work", the students benefited by having the opportunity of reflective thinking in Spanish. In the process of explaining the problem, the teacher used eighty-one Spanish words. Since the students saw and heard these words over and over again, they became quite familiar with them. However, the article failed to present evidence other than the author's opinion as to whether students learned Spanish better by this method than the traditional method.

Implications for My Teaching:

Although I do not feel personally qualified to teach by such a method at the present time, I believe it is an excellent method for those students who want to learn more than the basic elements of a foreign language. It requires that the teacher speak, read, write, and comprehend Spanish fluently. Until I read this article, the thought never occurred to me, at least consciously, of using the principles of reflective thinking in teaching a student to think in a foreign language. I was taught French by the grammar-translation method. For four years I have been an "A" student in French; yet I cannot converse with any fluency at all. The whole concept of "thinking" in another language is extremely difficult for me to understand. To have any fluency -- one must "think" in the language. When a student thinks in one language (English) and then must "translate" to another language (French, Spanish) before speaking -- he cannot develop much fluency. How can we pretend to "Teach for transfer" and still use a method of teaching that inhibits the transfer of foreign language skills to a "real-life situation? It looks to me as if this article may have a possible solution.

VII. Class Discussion and Activities

A. In general, the usual pattern for our two-hour meetings will consist of lecture or demonstrations the first hour and class discussion or activity the second hour. The exact plans will depend on the preferences of the individual instructor working with you.

B. There are two main purposes of the second hour activity:
   1. To give you an opportunity to integrate some of the ideas you have received from the readings, lectures and observations.
   2. To give you a chance to investigate some activities and techniques in which teachers need skill.

C. It is impossible to predict exactly what kind of activity will take place during the second hour. The specific content will depend upon your questions, comments, observations, and ideas. There will be many opportunities for you to use the particular talents you have. Here are some things which have been done in the past which may give you a general idea of what to expect:
   1. Committees may be set up to have the responsibility of preparing a research report (See Section VIII). Some of these committees may wish to volunteer or may be asked to act as resource persons and lead a class discussion on the topic they have prepared.
   2. Some individuals may wish to demonstrate a particular teaching technique to the class. You may, for example, wish to role-play a 9th grade teacher teaching his students (played by the rest of the class) some concept in algebra using a method which makes the problem seem real to the "students". Afterwards the class can make constructive suggestions and comments on how the method could be improved.
   3. Some of you may come across some especially stimulating book in the field of education or educational psychology and wish to lead the class in a short discussion of the ideas presented.
   4. The instructor may wish to demonstrate the uses that can be made of sociograms, sociodrama, etc. Class participation would be involved in the demonstration, and class discussion would follow.
   5. The instructor or any class member may propose some questions or ideas new to the field of education for class discussion. Discussion may relate some of the principles of learning developed in class to this new idea.
   6. Some individuals may wish to make observations of children at about the age level they expect to teach. Certain key observations may be reported to the class.
   7. Independent study and projects related to educational psychology may be allowed.

D. The above ideas are only illustrative of the type of activity that might occur. Obviously, the success of the second hour activity will depend on the quality of your participation and the responsibility you assume toward your intended profession.

VIII. SUGGESTIONS FOR PREPARING THE WRITTEN REPORT

Developing effective skills in writing term papers is a demanding process. To help you in your development the following suggestions have been prepared.

Choosing the Report Topic

Each student will choose a problem of interest to him within the first days of class. It is recommended that a topic be selected which has meaning to the student for his teaching field. The topic must relate to the subject matter of the course -- educational psychology.

Early in the quarter you will be asked to turn in an outline of your intended report. Your final report will be due about two weeks before the end of the quarter.

Topics

Broadly, the student will choose from the following areas for his report:

1. Learning
2. Individual Differences
3. Human Growth and Development
4. Measurement in Education
5. Personality

Preparing for the Report

As you do library reading, do the usual notemaking on five-by-eight cards. Use standard bibliographic procedure, listing the author, title, publisher, place of publication, date of publication and page of source.

Sources of Information

Of course, in the library the main sources of information on educational topics are:

1. The Encyclopedia of Educational Research
2. Education Index
3. The Card Catalogue
4. Reader's Guide to Periodical Literature
5. Psychological Abstracts

Writing the Report

A table of contents is essentially your outline of the paper. Most students have already learned that papers can be written much more easily if the table of contents is written first. The table of contents for the paper you are reading would look like this:

The Actual Writing

Content, not weight is the basis of your grade. Make your words count and do not pad. All evidence to date indicates that the appearance of the report does influence the reader, since his evaluation is a subjective one. Colorful papers impress us.

Titles. Use a system of titles and subtitles to make your organization clear by writing your outline into the body of the paper. The system we recommend is the one you are reading now. (Notice carefully how the concepts are broken down from major to minor ideas.)

References. Fully cite when credit is due. Paraphrase rather than extensively copy quotations into your paper.

Preferred System. There are several systems for citing references -- we recommend the following:

1. Use 5" x 8" cards to take notes.

2. Number your cards before you start from 1 to 50 (?).

3. Place the complete bibliographic information on a card before you begin taking notes.

4. Do not take notes from two or more references on one card. Use a separate card for each reference. If you take extensive notes from one reference so that more than one card is required, give the same number to the additional cards as appears on the first card.

5. When you write your rough draft, indicate your sources by placing the card number of each reference in parenthesis after the source's name. For example, you might say "... but a study by Jones (22) showed that ...."

6. If you wish to report the exact page where you found a certain idea or quotation, your citation might look like this: According to Lindgren, "Boys are more likely to fail annual promotions" (17, p. 160).

7. With this system no footnotes are necessary.

8. When you finish writing the rough draft, separate out those cards which you actually referred to in your rough draft.

9. Alphabetize those cards you actually used by author's last name.

10. Renumber those cards from 1 to 25 (?) (using red pencil) in alphabetical order.

11. In your rough draft go back and translate your old numbers to the corresponding red numbers.

12. When you type your list of references at the end of the paper, you will be able to type them in numerical and alphabetical order directly from the cards. Your list of references (if you only had 2) might look like this:

References

1. Coladarci, Arthur P., Educational Psychology. New York: Dryden Press, 1955.

2. Mowrer, O. H., "Motivation and Learning in Relation to the National Emergency", Psychological Bulletin, 1941, Vol. 38, 122-134.

13. Your end product then will enable a reader to identify the source of each idea or quotation by referring him to a certain numbered reference at the end of your paper. Your numbered references will also be in alphabetical order.

Presenting the Report

We fully expect that you will be proud of your report. We ask that it be typewritten and bound into a folder. No copies will be returned, so if you are certain you want a copy, invest in some carbon paper. If you are working on a committee report, bind all the reports from your committee into one folder. There will be one table of contents for the whole paper with chapters contributed by each committee member. Indicate which chapter you wrote.

IX. Field Trip

A. On May 20 we will have an opportunity to visit some school system and observe how teaching takes place from a new point of view.

B. There are three main purposes served by the field trip:
   1. You can see how the principles and concepts you have learn in class actually apply in the work of a teacher.
   2. You can begin shifting your point of view so that you no longer see the classroom through the eyes of a pupil but now see the classroom situation as a series of problems for the teacher to solve.
   3. You may get some fresh insight into your own vocational aspirations. What you see may help strengthen your resolve, or change your mind, on whether or not you want to become a teacher.

C. The cost of the trip is prorated among the class members. Last year the trip to Detroit cost each person $2.15 plus the cost of the lunch. The exact cost depends on how far we travel and how much you eat for lunch.

D. An extensive report of your experiences and insights is due on May 22 just 41 hours after your return. This relatively short period of time is deliberately set to encourage you to write your report while the details are still fresh in your mind. In the past students have found it wise to budget study and work time to allow a substantial block of time (6-8 hours) in which to write their report. Occasionally students find that other scheduled examinations or a heavy work load make it impossible to write a top quality report in the time allowed. If you have this problem, you may request from your instructor in advance of the trip permission to turn in your report on May 25 without penalty.

E. The field trip is an important part of the course. Consequently, you should make arrangements to get time off from your job if you work and make arrangements with any instructor with whom you have classes on Wednesday, May 20. About two weeks before the trip you will be given a written statement about the trip to present to your other instructors.

F. All details about the trip will be discussed in class on May 18. If you must be absent that day, be sure you get complete notes from someone else.

X. Examination Policy

A. The examinations are designed to help you as well as the instructors evaluate how well you are attaining the objectives of the course. They help determine your final grade which indicates the quality, not the quantity, of your achievements.

B. Short quizzes will be administered periodically over reading and lecture material. Since educational research indicates that distributed learning is generally better than massed learning, these quizzes are given to encourage you to study the assignments at the time they are given instead of waiting until the last minute to do all your studying.

C. No "make-ups" are allowed on quizzes. If you happen to be absent on the day a quiz is given, you have lost that score. However, since occasionally absences occur for legitimate reasons, it is our policy to scratch from your record the two lowest quiz scores you receive. Thus, if a person was never absent, the two lowest of his quiz scores would not be counted. If someone else was absent during two quizzes, then the two scores eliminated from his record would automatically be the two zeros from those days. Anyone having three or more excused absences which cause him to miss quizzes should see his instructor.

D. Anyone with an excused absence on the day of the mid-term examination, May 6, will be allowed to make it up on the following Thursday, May 9, from 3-5 p. m. See one of the instructors for exact location.

E. The final examination will last four hours and will occur probably on June 12. Some parts of the final examination will not affect your grade since you would answer these parts differently if this were the case. Still, in evaluating your progress it is important for us to know how you have changed some of your attitudes. Since we want honest answers, not the answers you think we want, these parts will not count toward your final grade.

[redacted]

[redacted] R64

[redacted] R25

[redacted] 192

[redacted] 4R3