The purpose of the study was to explore the effects of various practical instruction procedures on academic performance. Six sections of an introductory sociology course were assigned various experimental conditions involving quizzing and incentives. Examinations were administered at intervals during the course and once after course completion. Incentive conditions included grades, best-grade doubling, honor seminar, excused class attendance, and exempt final examinations. The major methodological problems confronted were variation in instructor behavior, within-group variation of student performance, comparative examinations, ineffective collection of study time data, and the brevity of periods to effectively manipulate experimental conditions. General findings were that two of four incentive conditions were effective in systematically modifying examination and study time performances, most of the information acquired in the course was lost fifteen weeks after course completion, neither the frequency nor scheduling of quizzes had a systematic effect on academic performance, and only a very low but significant relationship was found between study time and examination performance. (Author/CS)

FINAL REPORT

Project No. 7-C-024   PA 24

Contract No. OEG-1-7-07-0024-3900

LEARNING CONTINGENCIES IN THE COLLEGE CLASSROOM:

A PILOT STUDY

June 1968

U.S. DEPARTMENT OF

HEALTH, EDUCATION, AND WELFARE

OFFICE OF EDUCATION

BUREAU OF RESEARCH

FINAL REPORT

Project No. 7-C-024

Contract No. OEG-1-7-07-0024-3900

Learning Contingencies in The College Classroom:

A Pilot Study

James A. Wiggins - Hallowell Pope
University of North Carolina
Chapel Hill, North Carolina

Donald Bushell, Jr.
University of Kansas
Lawrence, Kansas

U.S. DEPARTMENT OF

HEALTH, EDUCATION, AND WELFARE

OFFICE OF EDUCATION

BUREAU OF RESEARCH

TABLE OF CONTENTS

Introductory Section

Summary.-- The purpose of the study was to explore the effects
of various practical instruction procedures on academic performance
(study patterns and examination performance). Six sections of an
introductory sociology course were assigned various experimental con-
ditions involving quizzing (both frequency and scheduling) and in-
centives (both positive and negative). Because randomization of
students was not possible, controls involved measurement of extraneous
variables (indicating only a single between-groups difference) and an
ABA-within-group design, Examinations were administered at course-
beginning, three periods during the course, course-completion, and
fifteen weeks after course-completion. Incentive conditions included
grades, best-grade doubling, honor seminar, excused class attendance,
and exempt final examinations.

The major methodological problems confronted were: (1) varia-
tion in instructor behavior; (2) within-group variation of student
performance; (3) comparative examinations; (4) ineffective collection
of study-time data; and (5) the brevity of periods to effectively
manipulate experimental conditions. The general findings were:
(1) two of four incentive conditions were effective in systematically
modifying examination and study-time performances (grades not being
one of them);(2) most of the information acquired in the course was
lost fifteen weeks after course-completion, regardless of experimental
conditions; (3) neither the frequency nor the scheduling of quizzes
had a systematic effect on academic performance; (4) only a very low,
but significant relationship was found between study time and exami-
nation performance.

Introduction.--Research on the effectiveness of classroom in-
struction and course procedures has up to now found no differences
among approaches, manifested inconsistent results, or demonstrated in-
creased effectiveness only with very complex, perhaps impractical, pro-
cedures. Our study searched for effective teaching procedures that
would be generally applicable and practical because they would not re-
quire basic institutional change or special training of classroom
instructors. We felt a promising beginning in this direction would be
the application of learning or "contingency" theory to the education
of the university student; the relative consistency of the experimental
studies of human learning and the few applications of these principles
to classroom learning have been encouraging.

Although one practical and much used application of the contin-
gency model of behavioral analysis has centered on programmed instruc-
tion, our research focused on applying this theory in other ways.
Because the classroom teacher has control over examination schedules,
grading weights for exams, and the rewards to be attached to grades
(other than intrinsic), we systematically used several variations in
these course procedures to test their impact. The course procedures
we tried were not innovative; rather we tried to investigate a variety
of procedures that are widely used in teaching college students. We
varied the weighting of quizzes, made movies and discussions as well as
exemption from classes and exams contingent on high quiz and exam per-
formance, varied the scheduling of quizzes, and finally, varied whether
or not quizzes were announced.

METHODOLOGY

Methods.--Six sections of the introductory sociology course at
the University of North Carolina at Chapel Hill received different
treatments of instruction. Because randomization of the students was
impractical, two other methods of controls were used: (1) Each section
was used as its own control in an ABA experimental design; (2) Measure-
ments of possible extraneous variables were made and statistically con-
trolled.

During the semester prior to the experimentation, several meet-
ings were held for the purpose of standardizing the behavior of the
instructors while in the classroom. Verbal reports from the instructors
provided the basis for meetings held to review the instructors' class-
room behavior and to encourage continued standardization. None of the
researchers who designed the experiment were also instructors in the
classroom. As will be suggested later, the success of the standardiza-
tion procedures is still in question.

Because our focus was on the effect of course procedures rather than on variations in presentation of course content, a common text was used in all sections (A. W. Gouldner and H. P. Gouldner, Modern Sociology, Harcourt, Brace & World, New York, 1963), and lectures were given to supplement the text, but did not deal with it directly. For all sections, each period of the semester was followed by an announced, one-hour examination (Exams I, II, and III)--each weighted equally for grading purposes. An outline of the text material to be covered in each hour exam was given each student at the beginning of the semester. The portion of this test over the text material provided our measure of course academic achievement. (Because we wanted the main emphasis on the text, the portion of the exam on the lectures was weighted at 15 per cent. The students in all sections were graded on fixed criteria (90 per cent correct or better on exams for an A, 80 per cent for a B, etc.), as opposed to a curve.

Although conducted on the same floor of the same building and approximately the same size, the sections were taught in different rooms, on different days of the week (Monday, Wednesday, Friday or Tuesday, Thursday, Saturday), and at different times of the day. The possible effects of these or related variables are controlled only by the ABA design of the experiment.

Randomization of students to treatments was impractical. Therefore, control relied on a relatively large sample size (and the measurement of possible extraneous variables. The first measurement was the mean student cumulative grade-point average prior to the experimentation. In almost all cases, this involved at least three semester's work. The second measurement was the mean student SAT-Verbal score. The third measurement was a 45 item "pretest" examination given the first day in class evaluating the student's prior knowledge of the course material. The items were taken from the test booklet for the course textbook. Test-anxiety was measured by the following items taken from the Alpert-Haber Test Anxiety scale.

1. Nervousness while taking an exam or test hinders me from doing well.

    Always _____ _____ _____ _____ _____ Never
           5   4   3   2   1

2. In a course where I have been doing poorly, my fear of a bad grade cuts down my efficiency on taking tests.

    Never _____ _____ _____ _____ _____ Always
          1   2   3   4   5

3

3. When I don't do well on a difficult item at the beginning
   of an exam, it tends to upset me so that I block on even
   easy questions later on.

   This never   This almost always
   happens to me ____ ____ ____ ____ ____   happens to me
                   1   2   3   4   5

The student's perception of his ability to affect the grades he receives
was measured by a revision of the education items from the Rotter-
Seeman-Liverant Internal vs. External Control scale.

1. Many times exam questions tend to be so unrelated to course
   work that studying is really useless.

   Strongly agree ____ ____ ____ ____ ____ Strongly
                   5   4   3   2   1   disagree

2. Sometimes I can't understand how teachers arrive at the
   grades they give.

   Strongly agree ____ ____ ____ ____ ____ Strongly
                   5   4   3   2   1   disagree

3. There is a direct connection between how hard I study and
   the grade I get.

   Strongly agree ____ ____ ____ ____ ____ Strongly
                   1   2   3   4   5   disagree

The student's concern for his academic performance was measured by the
following items:

1. Circle the number below which shows how important it is to
   you to do well in sociology.

   1 2 3 4 5   6 7 8 9 10

2. How often do you feel worried or bothered about what
   instructors think of you intellectually?

   Very often ____ ____ ____ ____ ____ Practically Never
               5   4   3   2   1

3. How often do you feel worried or bothered about what your
   friends think of you intellectually?

   Very often ____ ____ ____ ____ ____ Practically never
               5   4   3   2   1

4

4. How often do you feel worried or bothered about what your parents think of you intellectually?

Very often _____ _____ _____ _____ _____ Practically never
              5   4   3   2   1

5. Do you ever feel so discouraged with your self intellectually that you wonder whether anything is worthwhile?

This never   This almost always
happens to me _____ _____ _____ _____ _____ happens to me
               1   2   3   4   5

6. When you have made an embarrassing mistake or have done something that makes you look foolish in class or on an exam, how long do you usually keep on worrying about it?

A long time _____ _____ _____ _____ _____ No time at all
              5   4   3   2   1

Each of the above scales was administered both the first and last days of the semester and was coded by simply summing the numbers under the responses the subjects gave to each item. (All of the items had been pretested the previous semester and found to discriminate introductory sociology students.)

The students did not know of the experiment or which treatment they were to receive until they actually attended the class. The within-group comparisons of the ABA design of the experiment also provided some control over the contamination between sections.

The semester of 45 class meetings was divided into three periods (five weeks or 15 class meetings). At the beginning of the first period, all of the groups were given the following instructions:

> The Sociology Department of the University of North
> Carolina, in cooperation with departments of other
> schools, is investigating various procedures for pre-
> senting introductory sociology courses. For this
> reason, all participating classes will share a number
> of characteristics, including the textbook, Modern
> Sociology, by Gouldner and Gouldner.
>
> There will be no term paper required for the course,
> nor will there be any required reading outside the
> text. Thus, all the material which you will be
> examined on will come directly from the text and from
> the lectures. A standard grading criterion will be
> used rather than a competitive curve. Thus, each

student is graded on the basis of his performance relative to the assigned material, rather than on his performance relative to other students. A's will be given to all students whose examination and quiz scores average between 90% and 100%; B, 80-89%; C, 70-79%; D, 60-69%. The tests given will be made up primarily of objective-type questions.

Because the final grade will depend upon the performance average, it will not be necessary (indeed not possible) to make up any of the tests. There is no penalty for missing a test--this will only increase the relative importance of the remaining tests. Tests will cover all the assigned material presented up to that time, although they will be weighted toward the material introduced since the last test.

All of the groups were then instructed as to their meeting times and the dates and weights of examinations and quizzes.

Table 1 outlines the different course procedures in detail for 6 of the 8 groups (the other 2 will not be discussed in this paper). All quizzes counted toward the student's grade, with the exception of those advisory quizzes in the middle period of the semester (Period II) for Section 1. All quizzes were announced in advance except for the "pop" quizzes in the first and last periods of the semester (Periods I and III) for group 6.

The _positive-incentive group_ was given the following additional instructions:

To compensate for the fact that it is difficult to perform uniformly well on all occasions, you will be allowed to count your best three quizzes and your best exam double for grading purposes. All double scores, however, must be drawn from the first and last five-week periods of the semester. No quiz scores from the middle five-week period may be doubled, nor may the April 14 exam score (exam given at the end of the second five-week period) be doubled. In addition, during the first and last five-week periods, students achieving 90% or better on the quizzes and examinations will be eligible to attend an "Honor Seminar" held the week following each scheduled quiz or examination. These sessions will offer interesting film documentaries and discussions in supplement to the regular course content.

Table 1.--Course Procedures

Label for Treatment	Section (Treatment) Number	Course Procedures		
		Period	Number of Quizzes	Types of Quizzes
Regular Quizzes	3	I	0	
		II	4	Regular (count toward grade)
		III	0	
Advisory Quizzes- Grade Incentives	1	I	2	Regular
		II	2	Advisory (do not count toward grade)
		III	4	Regular
Unannounced Quizzes	6	I	2	Unannounced ("pop")
		II	4	Announced
		III	4	Unannounced ("pop")
Quizzes Plus Positive Incentives	2	I	2	Positive incentives attached
		II	2	No incentives
		III	4	Positive incentives attached
Quizzes Plus Negative Incentives	4	I	2	Negative incentives attached
		II	2	No incentives
		III	4	Negative incentives attached
Quizzes Plus Combined Incentives	8	I	2	Both positive and negative incentives attached
		II	2	No incentives
		III	4	Both positive and negative incentives attached

The <u>negative-incentive group</u> was given the following instructions:

Saturday classes will not be required for students who
achieve 90% or better on the scheduled quizzes during
the first and last five weeks of the semester. Satur-
day classes will be conducted as usual during the
middle five weeks of the semester. In addition, no
final examination will be required for those students
who maintain a 90% average during the first and last
five weeks of the semester.

The <u>combined-incentives group</u> was given both the "positive" and
the "negative" instructions.

At the beginning of the second five-week period, <u>the grade-incen-
tive (advisory quizzes) group</u> was given the following instructions:

During the next five weeks of the semester--the quiz
procedures will be changed slightly. The next two
quizzes--will not count toward your final grade. These
two quizzes will simply enable you to judge for your-
self how well you are mastering the course material and
thus provide a basis for preparing for the second
examination to be held on April 14. The quizzes will
be graded and discussed in the usual fashion, but the
grades will not be recorded--the results are for your
information only.

At the beginning of the last five-week period, this group was told that
the procedures would return to those existing during the first five
weeks except during the period there would be four quizzes instead of
two.

Examination performance was measured by three procedures in ad-
dition to the pretest examination. The first involved 40-item exams
administered at the end of the first and second periods and a 30-item
exam administered at the end of the last period. The second procedure
involved a 20-item exam given at the end of the last period and was
composed of items from the pretest examination given during the first
day of the semester. Third procedure involved a 35-item exam administer-
ed fifteen weeks after course-completion. The examination was composed
of the same 20 items used in the second procedure as well as five items
from each of the exams used in the first procedure.

A sample of twelve students, stratified by the cumulative results
of the 30- and 20-item exams given at course-completion, was drawn from
each of the groups. We will refer to the 20 items common to exams
administered at course-beginning, course-completion, and 15 weeks later
as the <u>related examination</u>. All others will be called <u>unrelated</u>
<u>examinations</u>. All data was analyzed in terms of per-cent item correct.

One difference among these various procedures, including the pretest examination was the consequence for performing each. The only consequence for performing the pretest exam was the assistance it gave to the researchers. The consequence for performing the remaining exams administered during the semester was the grade the student received. The consequence for performing the examination given fifteen weeks after course-completion was $4.00 given to those students who were in the upper quartile of the exam scores. The effects of these various consequences is of course unknown.

It is crucial for a within-group comparison of course performance between periods of the semester that we establish the comparable difficulty of the exams given at the end of each of these periods. Otherwise, differences can be attributed to variability in what the exams measured as well as to differences in student performance. The following procedures were employed to insure exam comparability as well as a standard response from the students to the exams throughout the semester: (1) the exam questions were of the objective, multiple-choice type; (2) they were written by the same person for the entire semester, following the same procedures; (3) the students were informed at the beginning of the course when the exams would be and what they would cover as well as their objective nature. The questions were written by a graduate student in sociology with prior graduate experience in education. She read the text carefully and provided a pool of objective questions that covered the assigned readings in the text, page by page. From this pool, several forms of each exam were constructed. Exams I and II had 40 items each and Exam III, 30 items--plus items common to the pretest. Our intent was to test for a thorough reading of the text by the student rather than to test abstract thinking or integrative ability. The similar standard deviations in the three exams for a given section plus the similarity in the overall means for the three exams increase our confidence that the exams were of equal difficulty.

In order to measure the students' study time, all sections were given the following instructions:

> In order to improve our understanding of how students distribute their study time, you are being asked to keep track of the number of minutes spent each day studying---. THESE RECORDS HAVE NOTHING TO DO WITH YOUR GRADE IN THIS COURSE. Indeed, the records will not even be processed by your instructor.

> "Study time" includes the time each day you spend reading your text, going over lecture notes, outlining the book, or even discussing the course material with another student.

> Please be candid. If you don't study a course for two weeks, enter two weeks worth of zeros on the record.

These records will only be of value if they are accurately
kept. For convenience, the records will be kept one week
at a time. Each week a completed record will be turned
in and a new one picked up to be filled out.

Thank you for your cooperation.

The records kept by the students were a single sheet of paper with
spaces for each day of the week in which the student could enter the
number of minutes he studied for the course on that particular day.

We had planned every week to collect a record of each student's
daily study time for the introductory course (as well as for two com-
parable courses), but because of administrative problems and the result-
ing mountain of paper and because we could not rely on eager student
cooperation, the project received only imcomplete study-time data. To
utilize the available data, we retained in the study-time analyses only
those students for whom we had data on at least three of the weeks for
a given period of the semester. On the basis of the available data, we
computed the mean weekly minutes of study time. One of the sections
gave almost complete data by this criterion (over four-fifths of the
students providing sufficient data in each of the three periods of the
semester), two supplied "adequate" data (three-fifths of the students
providing data in each of the periods), whereas in the remaining three
sections the data is not very adequate (less than one-half of the
students providing data in at least one of the three-semester periods).
Thus, we do not consider the results of any analysis involving study
time as definitive. However, we may use them with some confidence be-
cause in comparing students who did and did not report study-time data,
we find no difference between them on the Scholastic Aptitude Test
verbal scores or the course-relevant information with which they
entered the course (by t-test). However, the students reporting study-
time data had a higher mean cumulative grade-point average than those
who did not for the first two periods of the semester.

The student's attitudes toward various aspects of the course
were measured by various items on the questionnaire administered at
course-completion. The items asked for the student's evaluation of the
instructor's teaching skill, the course procedures, and the researchers.

All of the analysis of the data was made using the t-test.
Within-group comparisons were made using paired observations.

Findings and Analysis

This section will be divided into two parts for purposes
of analysis: (1) Effects of Incentive Conditions; (2) Effects of
Quizzing Procedures;

Effects of Incentive Conditions

Pretest: General Ability.--Two measures of general ability were
used as a basis of comparing the groups prior to the experiment--SAT-
Verbal Scores and Cumulative Grade-Point Averages. Table 2 shows the
results of this comparison. None of the between-group differences of

Table 2. Pretest: General Ability

	Group Mean Scores				Significance (p value) of Between-Group Differences					
	1	2	4	8	1-2	1-4	1-8	2-4	2-8	4-8
SAT-Verbal	540 [a] (40)	539 (43)	559 (44)	530 (34)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
CGPA	2.44	2.05	2.48	2.06	<.01	n.s.	<.05	<.001	n.s.	<.02

a   Number in parenthesis is N on which mean score was obtained.

the SAT-Verbal Scores were statistically significant. Such was not the
case concerning the between-group differences of the CGPA where the
averages of groups 1 and 4 were significantly greater than those of
groups 2 and 8. Thus, any conclusions concerning between-group differ-
ences in the dependent variables will have to be made somewhat con-
servatively and more reliance placed on the within-group differences.

Overall Semester Performance: Examinations.--Table 3 shows the
results of the analysis of the unrelated examinations and related
examinations (the latter being the 20 items used three different times).
Looking at the results of the pretest of both examinations, we see there
were no significant between-group differences. In other words, the
course-relevant information possessed by the students in the various
groups at the beginning of the course was relatively similar.

11

Table 3. Overall Semester Performance: Examinations

Unrelated Examinations	Group Mean Scores				Significance (p value) of Between-Group Differences					
	1	2	4	8	1-2	1-4	1-8	2-4	2-8	4-8
Course-Beginning (I)	50 [a] (43)	50 (51)	52 (42)	51 (40)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Course-Completion(II)	74 (44)	73 (51)	75 (45)	69 (42)	n.s.	<.05	<.01	n.s.	<.02	n.s.
15 Weeks Later (III)	53 (12)	54 (12)	57 (12)	52 (12)	n.s	n.s.	n.s.	n.s.	n.s.	n.s.
Diff: I-II	<.001	<.001	<.001	<.001						
Diff: I-III	n.s.	n.s.	n.s.	n.s.						
Diff: II-III	<.001	<.001	<.01	<.01						

Related Examinations	Group Mean Scores				Significance (p value) of Between-Group Differences					
	1	2	4	8	1-2	1-4	1-8	2-4	2-8	4-8
Course-Beginning (I)	23 (44)	21 (51)	22 (41)	24 (40)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Course-Completion(II)	61 (44)	64 (51)	58 (45)	59 (42)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
15 Weeks Later (III)	48 (12)	46 (12)	49 (12)	43 (12)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Diff: I-II	<.001	<.001	<.001	<.001						
Diff: I-III	<.001	<.001	<.001	<.01						
Diff: II-III	n.s.[b]	<.001	<.05	<.001						

a All scores are group means of individual per-cent correct scores.

b Non-significance is due to a lower mean course-completion score for the 12 students used in the examination 15 weeks later than for all 44 students which were present at course-completion. This was the only group where such a discrepancy occurred.

The results show there was a significant increase in the examination preformances on both examinations for all four groups from course-beginning to course-completion. Disappointingly enough, the results also show there was a significant decrease in the performances on both examinations for all groups fifteen weeks after the completion of the course. The results of the unrelated examinations show there were no significant differences between the performances at the beginning of the course and those fifteen weeks after course-completion. The results of the related examination show that although there was a significant decrease in performance from course-completion to a point fifteen weeks later, the latter performance was still significantly higher than the performances achieved at the beginning of the course.

The analysis of the between-group differences in the unrelated examination at course-completion found that only group 8 was significantly different from the other groups--in this case, significantly lower. A similar analysis of the related examination found no significant differences. No significant differences were found in either examinations fifteen weeks after course-completion.

Overall Semester Performance: Study Time.--Table 4 shows the average study time (in minutes) for the four groups. The only signifi--

Table 4. Overall Semester Performance: Study Time

	Group Mean Scores (minutes per week)				Significance (p value) of Between-Group Differences					
	1	2	4	8	1-2	1-4	1-8	2-4	2-8	4-8
Study Time Prior to Course	378 (33)	364 (36)	427 (41)	413 (32)	n.s.	n.s.	n.s.	<.05	n.s.	n.s.
Study Time During Course	283 (6)	257 (23)	338 (21)	298 (17)	n.s.	n.s.	n.s.	<.01	n.s.	n.s.

cant between-group difference found to be significant was between groups 2 and 4. The significance of this difference is probably minimized when one considers the fact there was a similar difference in the average weekly study time in previous courses reported by the students. Because of the tenuous reliability of the latter data, no within-group differences were examined.

Overall Semester Performance: Attitudes.--Table 5 shows the results of a questionnaire administered at the beginning and the end of the course. The results concerning the general measures of text-anxiety, internal-control, and academic concern show the following:

Table 5. Overall Semester Performance: Attitudes

	Group Mean Scores				Significance (p value) of Between-Group Differences					
	1	2	4	8	1-2	1-4	1-8	2-4	2-8	4-8
Pretest Attitudes	(37)	(39)	(43)	(33)						
Test-Anxiety	8.08	8.69	7.79	8.21	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
I.-E. Control	7.16	6.95	7.02	7.18	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Academic Concern	22.3	22.8	23.3	23.2	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Posttest Attitudes	(40)	(43)	(44)	(36)						
Test Anxiety	8.20	8.63	8.02	8.33	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
I.-E. Control	8.25	8.00	8.80	8.78	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Academic Concern	19.0	20.8	20.3	21.7	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Teacher Eval.	11.82	10.91	9.20	11.75	n.s.	<.001	n.s	<.05	n.s.	<.01
Procedure Eval.	4.82	6.47	4.91	5.58	<.01	n.s.	n.s.	<.001	n.s.	<.01
Researcher Eval.	4.65	5.88	4.11	6.58	<.05	n.s.	<.001	< .01	n.s.	<.001

(1) no significant between-group differences existed at the course-beginning; (2) no significant between-group differences existed at course-completion; and (3) no significant within-group differences occurred from course-beginning to course-completion. The results of the analysis of the students' evaluations of the instructor's skills, course procedures, and the researchers obtained at course-completion shows that all three objects were more negatively evaluated in group 4 than in any other group.

Within-Semester Performance: Examinations.--Table 6 shows the results of the examinations given at the end of each of the three five-week periods. (One should keep in mind that the examination given at the beginning of the course indicated the groups were nearly equivalent in course-relevant information.) Group 8 significantly decreased its performance from examination I to examination II as well as significantly increased its performance from examination II to III. Group 2 comes very close to replicating group 8 with the exception that the performance decrease from examination I to examination II is not quite statistically significant at the .05 level. Group 1 showed a significant increase in performance from examination II to III, but no decrease from examination I to II. Group 4 showed no significant changes.

14

Table 6. Within-Semester Performance: Examinations

	Group Mean Scores				Significance (p value) of Between-Group Differences					
	1 (44)	2 (43)	4 (45)	8 (39)	1-2	1-4	1-8	2-4	2-8	4-8
Condition A (Period 1)	71	73	75	72	n.s.	<.05	n.s.	n.s.	n.s.	n.s.
Condition B (Period 2)	70	70	74	65	n.s.	<.05	<.06	<.05	<.05	<.001
Condition A_2 (Period 3)	81	77	75	70	<.06	<.005	<.001	n.s.	<.01	n.s.

Diff: A-B	n.s.	<.07	n.s.	<.001
Diff: A-A_2	<.001	<.01	n.s.	n.s.
Diff: B-A_2	<.001	<.001	n.s.	<.05

The between-group differences were unsystematic. At the end of the first five-week period, the only significant difference existed where group 4 had a higher examination performance than group 1. At the end of the second five-week period, group 4 was higher than all other groups whereas group 8 was significantly lower than all other groups. At the end of the third five-week period, group 1 was significantly higher than groups 4 and 8 and close to being higher than all other groups. Group 8 was significantly lower than groups 1 and 2 and again close to being lower than all other groups.

Within-Semester Performance: Study Time.--The study time results shown in Table 7 are somewhat more consistent than the examination data. In three of four groups, the average study time per week decreased from the first period to the second and then increased from the second to the third. The only exception, group 4, maintained a high level of study time during the second period where the other groups had decreased their study time.

In contrast, the tremendous within-group variation in study time resulted in few significant between-group differences. The only statistically significant differences involved group 2 which was found to have a lower study-time rate than the other groups during the second period.

15

Table 7. Within-Semester Performance: Study Time

	Group Mean Scores (minutes per week)				Significance (p value) of Between-Group Differences					
	1	2	4	8	1-2	1-4	1-8	2-4	2-8	4-8
Condition A_1 (Period 1)	342	269	313	308	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Condition B (Period 2)	239	194	325	240	<.06 (15)[b]	n.s.	n.s.	<.001	<.02 (19)	n.s.
Condition A_2 (Period 3)	271	303	349	343	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

	1	2	4	8
Diff: A_1-B	<.02 (15)[a]	<.001 (27)	n.s.	<.02 (17)
Diff: A_1-A_2	n.s.	n.s.	n.s.	n.s.
Diff: B-A_2	n.s.	<.001 (25)	n.s.	<.02 (25)

a Number of students on which specific analysis was performed.

b Smallest-sized group of the two-group comparisons.

Effects of Quizzing Procedures

This section will consider the treatment variations in the number and type of quizzes and their impact on student-exam performance and mean-weekly study time. In closing this section, we will also look at the relationship between study time and exam performance (grades) by section as well as for individual students. It may surprise the reader that we find no simple effect of any of the following on student performance and study time: quiz frequency; a quiz's weight toward a grade; or whether a quiz is announced. Further, we find no more than a weak relationship between study time "put in" and exam scores.

Quizzes and Performance.--Table 8 gives the results for the mean exam scores for each group-period as well as percentage-point differences within each group and t-tests for the 3 possible comparisons within each group. Table 9 summarizes the results from the perspective of the prediction that the greater number of quizzes in a period preceding an exam will result in improved performance.

16

Table 8. Group Means and Standard Deviations Plus T-Tests for Exams I, II, and III.

Group and Quiz Condition[a]		Exam			Percentage Point Difference			T-Test (2 tailed)[c]		
		I	II	III	I-II	II-III	I-III	I-II	II-III	I-III
3 Regular Quizzes (0,4,0)	X̄	64.2	70.1	73.4	-5.9	-3.3	-9.2	** -3.456 (N=45)	* -2.114 (N=45)	** -6.653 (N=46)
	S	10.8	12.8	9.5						
	N[b]	46	45	46						
1 Advisory Quizzes (2,2,4)	X̄	70.6	69.5	80.9	1.1	-11.4	-10.3	0.617 (N=44)	** -7.448 (N=44)	** -6.793 (N=44)
	S	11.8	11.0	9.7						
	N	44	44	44						
2 Positive Incentives (2,2,4)	X̄	71.4	68.3	76.1	3.1	-7.8	-4.7	1.888 (N=43)	** -3.569 (N=47)	** -3.298 (N=47)
	S	10.3	12.6	10.7						
	N[b]	47	47	51						
4 Negative Incentives (2,2,4)	X̄	75.4	74.1	74.5	1.3	-0.3	0.9	0.706 (N=45)	-0.223 (N=45)	0.506 (N=45)
	S	10.0	10.3	11.0						
	N	45	45	45						
8 Combined Incentives (2,2,4)	X̄	72.2	65.4	70.1	6.8	-4.7	2.1	** 3.896 (N=39)	* -2.008 (N=40)	1.112 (N=41)
	S	11.1	10.4	13.2						
	N[b]	41	40	42						

Table 8 (continued)

Group and Quiz Condition[a]	Exam			Percentage Point Difference			T-Test (2 tailed)[c]		
	I	II	III	I-II	II-III	I-III	I-II	II-III	I-III
6 Accounced Quizzes (2,4,4,) X̄	70.1	71.7	71.4	-1.6	0.3	-1.3	-0.691 (N=31)	0.140 (N=31)	-0.667 (N=31)
S	11.6	11.5	9.3						
N	31	31	31						
TOTAL X̄	70.4	70.1	74.4	0.3	4.3	4.0	--	--	--

a   See text and Table 1 for description of course procedures. In parentheses the number of quizzes given in each period of the semester is recorded, e.g., (2,2,4) means that 2 quizzes were given in the first and second periods of the semester and 4 in the third period.

b   N's vary slightly because of student absences from exams

c   In this and subsequent tables all t-tests are for paired observations. See A. L. Edwards, Statistical Methods for the Behavioral Sciences, Rinehart, New York, 1954, pp. 278ff.

*.01 < p < .05

** p < .01

Table 9. Summary of Effects of Number of Quizzes on Mean Exam Scores

| Group Number (a) | Condition Label (b) | Quiz Pattern (c) | Total Comparisons Per Group (d) | Comparisons[a] Between Mean Exam Scores for Periods With: | | | | |
| | | | | Same Number of Quizzes | | Different Number of Quizzes | | |
				No Difference (e)	Difference (f)	No Difference (g)	Exam Score Higher in Period with More Quizzes (h)	Exam Score higher in Period with Fewer Quizzes (i)
3	Regular Quizzes	0,4,0,	3		1		1	1
1	Advisory Quizzes	2,2,4[b]	3	1			2	
2	Positive Incentives	2,2,4	3	1			2	
4	Negative Incentives	2,2,4	3	1		2		
8	Combined Incentives	2,2,4	3	1	1	1	1	
6	Unan-nounced Quizzes	2,4,4,	3	1		2		
Total Comparisons			18	4	2	5	6	1
Results as Predicted				Yes	No	No	Yes	No

Total Comparisons with Results as Predicted: 10[b]

Total Comparisons with Results not as Predicted: 8[b]

Table 9 (continued)

a   The mean exam scores for each of the three periods of the semester for a given group were compared with each other, giving 3 comparisons per group and 18 comparisons in all. We expected that the more quizzes in a period, the higher the mean exam score. Thus, the following results for mean exam scores are contrary to predictions: (1) any difference between two periods for which the number of quizzes are the same; (2) no difference between two periods for which the number of quizzes are different; and (3) a lower score for periods that had fewer quizzes than another period.

b   Period II for group 1 had quizzes which did not count toward the student's grade. If these are not considered as quizzes (no external consequences are contingent on quiz performance), this would lower the number of predicted results from 10 to 9 and up the number of non-predicted results from 8 to 9.

There are differences in performances among periods of the semester, 8 of 18 possible comparisons being statistically significant at the 5 per-cent level or better. However, it appears that no single, simple interpretive scheme will account for the results. Frequency of quizzes does not consistently result in improved performance or in decreased dispersion of performance scores.

That quizzes do not have the expected impact is illustrated in the regular quiz treatment (a BAB design). Here no quizzes during the first period resulted in a low performance level (6 percentage points below the mean for all students) and the introduction of 4 quizzes during the second third of the semester resulted in a significant improvement in performance. During the last third of the semester, we anticipated a decrease in performance approximating the first period's performance--again, no quizzes were given in this section. However, the performance on Exam III was a step above that of Exam II (by 3.3 percentage points, $p < .05$). Quizzes were not needed to produce continued improvement, even though Exam III counted no more toward the student's final grade than did Exams I or II. All of us can supply post-hoc explanations. One possibility: having been "punished" by low scores (grades) on Exam I and "forced" to improve by quizzes on Exam II, the students continued to act on the lesson learned in Period II during Period III and studied "hard" and so did well on Exam III (their mean-weekly study time did increase in Period III). But we would urge the reader to forego such speculations until the results of more controlled experiments are reported.

The regular-quiz group provided the only example of higher performance, even though the number of quizzes was less, and it also provided the only example of performance increasing (by 9.2 percentage points) with the number of quizzes remaining constant. However, there is also an instance in the combined incentive's condition of performance decreasing (by 6.8 percentage points) even though the number of quizzes remained the same. Our project has the dubious distinction of producing the full range of results, expected and otherwise, in observing student response to variations in numbers of quizzes.

Table 9 provides a way of summarizing our results. We, of course, consider a difference in mean performance score to be a meaningful difference only if it reaches statistical significance (at the 5 per-cent level). We can see that for 6 comparisons, the number of quizzes remained the same. As we have just discussed, there was 1 case each of increase and decrease in performance; whereas 4 comparisons show no difference in mean-exam scores within groups as predicted. These results are not terribly encouraging. When we consider the results for comparisons in which there were different numbers of quizzes, we are, perhaps, even less encouraged that we have a firm grasp of student behavior. Of 12 possible comparisons involving different numbers of quizzes, the following are the results: (1) 6 comparisons show a higher performance score in the period with most quizzes--as

21

expected; (2) 5 comparisons show no differences in exam performance--
quizzes had no impact; and (3) 1 comparison shows a decrease (true,
it is only a moderate decrease) in mean exam performance, even though
the number of quizzes was higher (0 as compared with 4). In sum, this
means that 10 of 18 comparisons gave results as expected. Note that
we have considered the advisory quizzes in Period II for group 1 as
real quizzes, even though no external consequence to the student,
such as a grade, was contingent on his quiz performance. If these
advisory quizzes are considered non-quizzes, we would have one less
result as predicted, because this group did not do worse in the
second than in the first period on exam performance.

Quizzes and Mean Weekly Study Time.--Let's turn to the results
that relate the number of quizzes per period to the averages of the
mean-weekly study times for all students in a group who reported
sufficient data in a period of the semester. The total study time
reported was divided by the number of weeks reported per period and
this figure was used as the average-weekly study time for a student
for a given third of the semester. All students who did not report
at least 3 weeks of study time are excluded from this analysis.
(Note that the N's in Tables 10 and 11 differ. Table 10 means in-
clude all students included in the analysis for a given period.
Table 11 included only those who reported data for both the periods
that are being compared. Therefore, the figures vary between the
two tables, but the conclusions drawn from each set of data are sub-
stantially the same.) For all students in all groups combined, the
mean-weekly study time in hours is 4.5, 4.1, and 5.1 for each period
of the semester (see Table 4). But note the standard deviations
around each of these means: 1.95, 1.95, and 2.23 hours respective-
ly. In any group-period the standard deviation never falls below 1.00
hours and it reaches 2.93 hours for the advisory-quiz condition in the
first period. As every college instructor knows, there is a high de-
gree of variability in how much students study. However, in our
study, the number of quizzes seems to have no systematic relation-
ship to study-time variability. Our experience indicates that a
rather large proportion of students are spending small amounts of
time studying and still passing sociology courses; on the other hand,
certain students are putting in what might be considered excessive
amounts of time (over 6-7 hours a week)--and some of them may be learn-
ing rather little. This possibility is discussed further when we re-
port the relationship between study time and performance.

Our results show no tendency for number of quizzes to produce
increased amount of study time during the period the quizzes are
given (see Table 11). We may summarize the results in a fashion simi-
lar to that for exam performance. (1) There are 6 comparisons be-
tween periods and within groups in which the number of quizzes if the
same. In 3 of these the mean-weekly study time increased and in 3 it
decreased. (2) There are 12 comparisons in which the number of quizzes
differed between the periods being compared. In none of these

Table 10. Minutes of Study Time Per Week

Group Number	Condition Label	Quiz Pattern[a]		Period I	Period II	Period III	Difference I-II	Difference II-III	Difference I-III
3	Regular Quizzes	0,4,0	$\bar{X}$ S N	222 92	253 94	261 127	-31	-8	-39
1	Advisory Quizzes	2,2,4	$\bar{X}$ S N	317 176 25	239 115 15	266 136 18	78	-27	51
2	Positive Incentives	2,2,4	$\bar{X}$ S N	269 84 31	176 97 35	307 114 36	93	-131	-38
4	Negative Incentives	2,2,4	$\bar{X}$ S N	305 103 41	314 150 27	346 146 33	-9	-32	-41
8	Combined Quizzes	2,4,4	$\bar{X}$ S N	307 116 32	243 83 19	340 118 27	54	-97	-33
6	Unan- nounced Quizzes	2,4,4,	$\bar{X}$ S N	174 61 15	191 111 11	204 134 8	-17	-13	-30
	TOTAL		$\bar{X}$ S N	270 117 249	245 116 213	306 134 213			
				(270= 4.5 hours)	(245= 4.1 hours)	(306= 5.1 hours			

a The number of quizzes given in Period I, II, and III of the semester.

Table 11. T-Tests On Mean Study Time For Students Reporting For The
Necessary Two Periods of The Semester

Group Number	Condition Label	Quiz Pattern[a]		Difference between Periods in Mean Minutes of Study Time and t-Values		
				I-II	II-III	I-III
3	Regular Quizzes	0,4,0	Difference between means (in minutes)	-30	0	-30
			t (2 tailed)	-2.342	.004	-1.760
			significance level	<.05	n.s.	n.s.
			N	43	36	37
1	Advisory Quizzes	2,2,4	Difference between means (in minutes)	103	-86	71
			t (2 tailed)	2.847	-2.140	1.437
			significance level	<.05	n.s.	n.s.
			N	15	6	11
2	Positive Incentives	2,2,4	Difference between means (in minutes)	81	-104	-31
			t (2 tailed)	4.271	-4.438	-2.057
			significance level	< .01	<.01	n.s.
			N	27	27	25
4	Negative Incentives	2,2,4	Difference between means (in minutes)	-10	-22	-36
			t (2 tailed)	-0.335	-0.807	-1.534
			significance level	n.s.	n.s.	n.s.
			N	25	22	32
8	Combined Incentives	2,2,4	Difference between means (in minutes)	65	-100	-32
			t (2 tailed)	2.712	-3.802	-1.660
			significance level	< .05	< .01	n.s.
			N	17	19	25
6	Unan- nounced Quizzes	2,4,4	Difference between means (in minutes)	-29	-40	-2
			t (2 tailed)	-0.741	-0.934	-.031
			significance level	n.s.	n.s.	n.s.
			N	9	5	7

a The t-tests are for paired observations and so only those students are
included who reported data for both periods being compared. Thus, the
N's and other figures vary between this table and Table 10.

24

comparisons was the mean-weekly study time higher in the period with
the fewer quizzes. However, in 9 of these comparisons there was no
statistically significant difference between the periods with differing
numbers of quizzes, whereas in only 3 of them did the period with the
greater number of quizzes have a greater mean-weekly study time--it
ranged from ½ hour to over 2 hours greater. By strict criterion
(statistical significance at the 5 per-cent level) only 6 of 18 com-
parisons turn out according to predictions, that is, with the greater
number of quizzes resulting in higher mean-weekly study time and no
difference in numbers of quizzes resulting in no study-time difference.

The reader should be reminded at this point of one unavoidable
consequence of increasing quiz frequency and yet retaining the same
number of hour exams and the same weighting of them for grading pur-
poses (for all groups each hour exam was weighted 20 or 25 per cent).
The unavoidable consequence is that this procedure at the same time
reduces the grade weighting given to each quiz. Our study design
dictated that the criterion measures of performance be given as nearly
equal grade weights as possible and this meant that the number of
quizzes could not be varied independently of quiz weights toward the
final grade. In the period with 2 quizzes, for example, the student
is faced with double the weight toward a grade for each quiz in com-
parison to the period with 4 quizzes.

Mean-Weekly Study Time and Performance.--We have reported the
impact of number of quizzes on performance and study time; now we will
briefly cover the association between study time and performance, both
for group-periods and individuals. We all assume that the student per-
ceives increased study time as resulting in improved grades and that is
why he will respond to increased numbers of quizzes with more studying--
if grades are contingent upon quiz performance. What do our results
indicate about study time and performance? First, let's look at the
rank order of group-periods (N=18) by the mean-weekly study time of
their students (based on Table 10) in comparison to the rank order of
these 18 group-periods by mean-exam performance (based on Table 8).
The rank-correlation coefficient for the rank order pairs is only .51
(p <.05), and there are some surprisingly large discrepancies in rank.
The advisory-quiz condition, Period III (4 regular quizzes) is first
in rank on exam performance, but only ninth in rank on mean study time.
The combined-incentives condition, Period III is second in rank on
study time, but no better than thirteenth in rank on exam performance.
The relationship between study time and performance based on group-
period comparisons is no better than moderate and indicates the de-
sirability of looking at this relationship among individual students--
although up to now we have not presented any analyses for individuals.

Table 12. Examination Performance and Study Time

		Percent Correct						
		<59	60-69	70-79	80-89	90-100	Total	Percent
		Exam I (given at end of week 5)						
Period I:	≤ 2	3	7	0	1	0	11	6
Mean Hours of	2-3	8	6	9	5	1	29	16
Study Time	3-4	3	12	14	6	1	36	20
Per Week	4-5	7	13	19	8	1	48	26
(weeks 1-5)	5-6	1	9	8	9	0	27	15
	6-7	1	4	7	6	2	20	11
	7-8	0	1	0	2	1	4	2
	> 8	2	0	1	4	2	9	5
	Total	25	52	58	41	8	184	
	Percent	14	28	32	22	4		100
		Exam II (given at end of week 11)						
Period II:	≤ 2	3	9	6	3	1	22	15
Mean Hours of	2-3	8	5	13	6	0	32	22
Study Time	3-4	3	7	8	8	0	26	18
Per Week	4-5	2	8	9	6	0	25	17
(weeks 6-10)	5-6	5	5	10	4	2	26	18
	6-7	2	1	2	5	1	11	7
	7-8	0	0	1	1	1	3	2
	> 8	0	0	2	1	0	3	2
	Total	23	35	51	34	5	148	
	Percent	16	24	34	23	3		100
		Exam III (Given at end of week 16)						
Period III:	≤ 2	0	2	6	6	0	14	9
Mean Hours of	2-3	1	7	5	0	1	14	9
Study Time	3-4	5	2	7	5	4	23	14
Per Week	4-5	2	8	16	6	3	35	22
(weeks 11-15)	5-6	4	2	11	8	2	27	17
	6-7	2	6	2	7	2	19	12
	7-8	0	1	3	4	6	14	9
	> 8	1	2	4	1	5	13	8
	Total	15	30	54	37	23	159	
	Percent	9	19	34	23	14		100

$Tau_c = .23$

$p < .001$

$Tau_c = .14$

$p < .01$

$Tau_c = .15$

$p < .01$

Table 12 presents the surprising results for cross-tabulations
of mean hours of study time per week by per cent correct on each of the
three-hour exams given during the semester. The measure of association
used (tau_c) indicates a weak relationship for each of the exams, even
though they each could have occurred by chance less than one time out
of a hundred. Can we expect no better relationship than this for
study time and performance? Does not studying result in more learning
and better test scores? What do we make of the students who studied
more than 5 hours on the average a week and yet scored below 60 per-
cent ("F"): 4 on Exam I, 7 on Exam II, and 7 again on Exam III? Note,
however, that studying little and scoring above 90 per cent was not
frequent: only 1 student who averaged less than 2 hours a week for a
period beat this mark on any of the exams given at the end of the
semester periods.

The obvious check is to see whether the "brighter" students can
make their studying pay off in grades, whereas the duller ones cannot.
We checked this dichotomizing the students on the basis of their
Scholastic Aptitude Test verbal scores, then their cumulative grade-
point averages upon entering the introductory-sociology course (based
usually on at least 3 prior semesters' work), and finally on their
course-relevant information upon entering the course (based on a pre-
exam of 20 objective test items). In comparison to the total group,
the relationship between study time and performance almost as often de-
creased as it increased among the higher-ability students when cate-
gorized by these variables for the 3 exams. (For the high SAT-verbal
and high information scorers on Exam II, the relationship was not sig-
nificantly different from zero--$p > .05$.)

One control variable did consistently change the association
between study time and performance: internal-external control situa-
tion. The questions indexing this variable essentially asked the stu-
dent for his perception of the connection between study time and grade
pay-off. Among those students feeling there was limited connection,
the relationship between study time and performance never reached sta-
tistical significance (tau_c's of $\pm.09$, $\pm.13$, and $\pm.02$ on the 3 exams
respectively). Among those students seeing a closer connection, the
relationship was weak but significant (always p .01 with tau_c's of
$\pm.26$, $\pm.21$, and $\pm.22$ respectively on the 3 exams). We have no way of
knowing whether those results indicate students have an accurate per-
ception of how things are for them or whether some have a tendency to
rationalize for inefficiently spent time. In any event, study time
and exam performance were only weakly related among this group of
students taking the introductory-sociology course.

Summary and Conclusions

Methodological Problems

1. Initial randomization of student assignment to experimental groups was not possible.

2. All of the pretests used in the study (SAT-Verbal, examination performance, test-anxiety, etc.) showed consistent results across all groups with the exception of the Cumulative Grade-Point Average.

3. Although reading material and course-procedure instructions were standardized, the lectures and classroom behavior of the instructors was not. Only groups 4 and 8 had the same instructor. Even if instructors themselves could have been initially "standardized" it would have been very difficult to assure the continuation of standardization because of the instructors' reaction to the changes resulting from the manipulations. To control this would require constant socialization of instructors throughout the semester, through in-class transmitters if possible.

4. A tremendous amount of within-group variation existed, particularly in groups 1 and 4. The above four problems are particularly relevant to the between-group comparisons.

5. The related examination had the advantage of allowing exact comparisons at three different points in time, but it also had the disadvantage of "testing effects" with the students having increasing experience with it.

6. The unrelated examinations had the opposite advantage and disadvantage of the related examination.

7. The examinations at the end of each five-week period had neither advantage. However, a trained person did attempt to prepare equivalent examinations and this person had minimal awareness of the condition changes.

8. The procedure for obtaining data on study time was inefficient and as a result much of the data was incomplete. If the researchers had taken the responsibility of collecting the weekly records rather than having relied on the different instructors, data-collection would probably have been improved. In addition, a contingency device to use on the students would have been useful.

9. Finally, the ABA design itself may have produced effects resulting from switching from one condition to another. There is no way to handle this problem other than by extending the length of each "A" in "B" period until change-effects have diminished or by repeating

28

the "A" and "B" periods enough times to achieve consistency across all "A" periods and across all "B" periods. Either procedure is impractical during a single semester. The last five problems are particularly relevant to within-group comparisons.

Incentive Conditions

1. Groups 2 and 8 (positive and combined incentives) conformed best to the predictions. In group 8, examination and study-time performances were significantly higher during both combined-incentive-present conditions than during the grade-incentive-only condition. Group 2 did the same with the exception of the switch from the first positive-incentive-present condition to the grade-only condition where significance was closely approached but not achieved on the examination performance data.

2. The data on group 4 (negative incentive) suggested that the lack of within-group change resulted from the relatively high performance (consistent in both examination and study time) during the negative-incentive-absent condition. This was also the group in which the instructor's skills, course-procedures, and the researchers were most negatively evaluated by the students. The study-time performance of group 1 approximated the expected ABA pattern, but the examination performance more closely approximated the usual late-semester cramming behavior on the part of students. The rationale for these findings is not obvious. Both of these groups had higher within-group variation on both examination and study-time performances, and both had higher Cumulative Grade-Point Averages than groups 2 and 8. An analysis dicotomizing the students into high and low CGPA classes found that the different incentive condition did not have a differential effect on examination performances of the two classes.

3. The most disappointing finding in this study has been the lack of significant within-group differences between the unrelated examinations given at the beginning of the course and given fifteen weeks after course-completion. Although the related examination too showed a significant decrease in performance from course-completion to fifteen weeks later, the later performance was still significantly higher than that at course-beginning. However, the students had had at least two previous experiences with the items in this examination by that time, so the results may indicate the effects of repetition. This would suggest that if the information in the course had been taught in such a way as to anticipate that information confronting the students onee outside the course (associating the two more closely), then the information may have been maintained longer due to repetition and association. In total, however, the results do indicate that most of the information acquired in the course was lost regardless of the various conditions existing

29

in the course. Of course, it cannot be concluded that nothing was maintained by the student. They may have changed their attitudes or may have acquired a different way of thinking. However, the goal of most introductory-sociology courses is primarily information-learning and this is exactly where the experimental conditions failed in the long run.

4. At least one major question is left unanswered, and it must remain so. What would have been the effect of contingency conditions as opposed to just incentive conditions? If more of the students had experienced the rewards for performance instead of just the promise of these rewards, would their performance have continued to be high and would others have emulated them? In fact, does the aspiration for a reward without most students being able to achieve it decrease the performance more than if there had been no rewards at all?

Quizzing Procedures

1. Neither the frequency nor the scheduling of quizzes had a systematic effect on study time or examination performance.

2. Only a very low but significant relationship existed between study time and examination performance. The only control variable affecting this relationship was "internal-external control."

(TOP)

ERIC ACCESSION NO.					
CLEARINGHOUSE ACCESSION NUMBER	RESUME DATE 6-15-68	P.A.	T.A.	IS DOCUMENT COPYRIGHTED? YES ☐ NO ☒	
				ERIC REPRODUCTION RELEASE? YES ☒ NO ☐	

001

100
101
102
103

TITLE
Learning Contingencies in the College Classroom: A Pilot Study Final Report

200

PERSONAL AUTHOR(S)
Wiggins, James A., and others

300

INSTITUTION (SOURCE)	SOURCE CODE
University of North Carolina, Chapel Hill, N. C., Dept. of Sociology	

310

REPORT/SERIES NO.

320

OTHER SOURCE	SOURCE CODE

330

OTHER REPORT NO.

340
350

OTHER SOURCE	SOURCE CODE

400

OTHER REPORT NO.

PUB'L. DATE 6-15-68	CONTRACT/GRANT NUMBER OEG-1-7-07-0024-3900

PAGINATION, ETC.

500
501

600
601
602
603
604
605
606

RETRIEVAL TERMS
College teaching, exam performance, study time, incentives, number of exams, longitudinal performance

607

IDENTIFIERS

800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822

ABSTRACT

The purpose of the study was to explore the effects of various practical instruction procedures on academic performance (study patterns and examination performance). Six sections of an introductory sociology course were assigned various experimental conditions involving quizzing (both frequency and scheduling) and incentives (both positive and negative). [Because randomization of students was not possible, controls involved measurement of extraneous variables (indicating only a single between-groups difference) and an ABA-within-group design.] Examinations were administered at course-beginning, three periods during the course, course-completion, and fifteen weeks after course-completion. Incentive conditions included grades, best-grade doubling, honor seminar, excused class attendance, and exempt final examinations.

The major methodological problems confronted were: (1) variation in instructor behavior, (2) within-group variation of student performance, (3) comparative examinations, (4) ineffective collection of study time data, and (5) the brievity of periods to effectively manipulate experimental conditions. The general findings were: (1) Two of four incentive conditions were effective in systematically modifying examination and study time performances (grades not being one of them); (2) Most of the information acquired in the course was lost fifteen weeks after course-completion, regardless of experimental conditions; (3) Neither the frequency nor the scheduling of quizzes had a systematic effect on academic performance; (4) Only a very low but significant relationship was found between study time and examination performance.