DOCUMENT RESUME

ED 100 360 IR 001 459

AUTHOR Ott, Mary Diederich
TITLE The Evaluation of Two Methods of College Physics Instruction. Final Report.
INSTITUTION Cornell Univ., Ithaca, N.Y. Center for Improvement of Undergraduate Education.
SPONS AGENCY Alfred P. Sloan Foundation, New York, N.Y.
PUB DATE Dec 74
NOTE 102p.

EDRS PRICE MF-$0.75 HC-$5.40 PLUS POSTAGE
DESCRIPTORS *Academic Achievement; Classroom Observation Techniques; College Science; Comparative Analysis; *Conventional Instruction; Course Evaluation; Higher Education; *Learning Laboratories; Lecture; *Participant Satisfaction; *Physics Instruction; Student Attitudes; Teaching Methods
IDENTIFIERS *Audio Tutorial Instruction; Mathematical Aptitude

ABSTRACT

Two instructional methods for a freshman physics course were compared: audio-tutorial (AT) instruction in a learning center and lecture-recitation-laboratory (LRL) instruction. Both random assignment and student preference were used to distribute the 575 students between the two methods. Course content, homework, and tests were identical. Achievement was evaluated by exam grades, adjusted for SAT math scores and math pretest scores. Questionnaires were used to gauge student attitudes toward the course. To determine long range effects of the rwo instructional methods, an analysis was made of the progress of the students, in terms of choice of physics classes, was designed to measure content and behavior under AT and LRL instruction, to show variations between the two methods, and to determine interrelationships among student achievement, student evaluation of teachers, and teacher behavior. Results indicated that AT and LRL lead to similar levels of achievement, but students with low math aptitude did better under AT while high math achievers did better under LRL. The observation instrument is included. (SL)

ED 110360

FINAL REPORT TO THE SLOAN FOUNDATION

The Evaluation of Two Methods
of College Physics Instruction

Mary Diederich Ott
Principal Investigator

December 1974
Center for the Improvement of Undergraduate Education
Cornell University
Ithaca, New York

Table of Contents

I. Introduction... 1

II. Achievement and Attitudes of Students in the Two Methods..... 5

A. Student Achievement..................................... 5

B. Student Attitudes......................................16

III. Longitudinal Study of Student Achievement....................20

IV. Observations of Instruction.................................25

A. Learning Center Observations............................25

1. Student Activities..................................27

2. Tutor Activities....................................30

3. Differences Among Teachers..........................30

4. Number of Student Hours in Learning Center...........32

B. Observations of Lecture and Recitation Instruction.......34

1. Lecture Observations and Recitation Observations.....35

2. Observation of Recitation Instruction: Method, Observer, and Teacher Effects........................40

a. Content Categories..............................42
b. Behavior Categories.............................44

3. Relations of Recitation Observations to Student Achievement.................................47

4. Relations of Recitation Observations to Teacher Ratings.....................................52

V. Conclusion: Guidelines for Course Decisions.................55

VI. Appendices

I. Introduction

This report is concerned with the evaluation of two methods of instruction used in a college physics course at Cornell University. The two methods are audio-tutorial instruction and lecture-recitation-laboratory instruction. The evaluation was conducted during the academic year 1973-74.

The research program investigated a number of related areas:

1. The effects of the two methods of instruction on levels of student achievement and on student attitudes toward the course.

2. Longer term effects of the two methods in terms of achievement and areas of interest subsequent to the course.

3. Characteristics of the major components of the two methods of instruction, i.e., lecture, recitation, and learning center instruction.

Following a description of the course, the methods of instruction, and the methods of assigning students, these three areas will be considered in detail. The report concludes with guidelines for the use of audio-tutorial instruction and lecture-recitation-laboratory instruction.

A number of persons at Cornell University have made significant contributions to this research. In particular, David Macklin of the Center for Improvement of Undergraduate Education has contributed to every phase of the work. Stephen Brock, also of the Center for Improvement of Undergraduate Education, served as project director. Walter Federer and Jason Millman provided advice on experimental design and statistical analysis. Donald Dickason, Donald Burgett, and Barbara Hirshfeld assisted by providing student aptitude test information. Patricia Musick, Catherine Clark, and William Lovejoy performed invaluable service as observers of instruction. The Division of Basic Studies of the College of Engineering, and the faculty, teaching assistants, and students in Physics 112 cooperated with the research project despite the occasional inconvenience it entailed.

The Course

The course in question, P112, Mechanics and Heat, is a one-semester freshman level physics course offered to approximately five hundred and fifty five engineering and physics majors each spring semester. In 1974 97% of the students were freshmen, 91% were males, and 90% or more were enrolled in engineering.

The subject matter of the course included classical mechanics, special relativity and heat. The course textbook was Fundamentals of Physics by David Halliday and Robert Resnick. This text was supplemented by notes on relativity written by the course professor, Jay Orear.

The Methods of Instruction

The course was offered to students via one of two methods or treatments -- a lecture-recitation-laboratory (standard) method of instruction, and an audio-tutorial (a.t.) method. The content of the instruction was similar in the two methods. In addition, students in both had the same homework assignments, the same lab assignments, and identical quizzes and examinations. Therefore the difference in treatments was primarily one of instructional method, rather than of content, assignments, or examinations.

The standard method of instruction included two hours of lecture and two hours of recitation per week, and a two-hour laboratory every other week. (Lecture and recitation are described in more detail in part IV E.) On the average, twenty six students were assigned to each recitation section and twelve to each lab section. Recitations and laboratories were taught by faculty members and graduate teaching assistants.

The audio-tutorial method included one hour of recitation per week. This hour of group instruction was included primarily to provide an opportunity for group interaction, as well as student contact with one particular instructor and an opportunity for testing. In the audio-tutorial method, all other instruction took place at the student's convenience in a learning center. The learning center was staffed by tutors forty-seven hours per week. Materials available in the learning center included demonstration apparatus for self-demonstrations, laboratory equipment, audio-tape commentaries, and slides. The taped commentaries and slides were coordinated with a study-guide developed by the author and Professor John Silcox. Audio tape-slide-study guide sequences were intended in particular to assist students in learning to solve physics problems. (The students' usage of the learning center is described in Part IVA.)

Methods of Assigning Students to Treatments

The research involved an experimental design intended to answer questions related to:

1. the methods of instruction (audio-tutorial or standard)
2. the procedures of assigning students to methods of instruction.

The two methods have already been described. The two procedures of assigning individual students to these methods were (a) random assignment, (b) assignment according to student preference.

Thus our experimental design can be diagrammed as follows:

	Audio-tutorial	Standard
Random		
Preference		

In detail, the students were assigned as follows:

Each enrolled student was assigned to one of twenty-two recitation sections. Students were assigned to sections at a given class hour on the basis of their other scheduled classes and their personal preferences. The fifteen sections scheduled at four of the class hours were then selected for inclusion in the study. (These four class hours were selected because three or more sections were scheduled at each of these hours.) Eight of these sections became standard recitation sections. Seven became audio-tutorial recitation sections.

At two of the four class hours included in the study, students were randomly assigned to treatments and to recitation sections within the treatments. At the other two hours, students were assigned to treatments on the basis of their preferences as expressed on a questionnaire administered at class meetings during the first week of the semester. Within each of these two hours, students were then randomly assigned to specific sections within the preferred treatment.* In 1974 the result of these procedures was the formation of four audio-tutorial random (ATR) sections, four standard random (STR) sections, three audio-tutorial preference (ATP) sections, and four standard preference (STP) sections, as indicated in the following diagram.

	Audio-tutorial	Standard
Random	4	4
Preference	3	4

Eight teachers, including two faculty members and six graduate students, taught these fifteen sections. Seven teachers taught one audio-tutorial and one standard section each. One graduate student taught two standard sections, only one of which was included in the study.

We turn now to consideration of the three areas of research referred to in the Introduction.

* At each of these hours, some students were assigned to a method contrary to their preference in order to balance class sizes. These 23 students are not included in the subsequent analyses.

II. Achievement and Attitudes of Students in the Two Methods

A. Student Achievement

In considering the achievement of the students, we conceptualized the problem as one of possible trait-treatment interactions. That is, we determined whether student characteristics such as aptitudes or previous achievement predicted greater achievement in one method (treatment) than in the other. The results of research completed early in this project indicated that in f such an interaction existed when the course was offered in the Spring semester, 1973. In that case, two "traits" were important -- achievement in math, and math aptitude. Stated briefly, the 1973 results were as follows.

Students with very high math aptitude (SAT Math scores of 725 or higher) and high math achievement on a course-specific pretest, had higher predicted grades in the standard method than did comparable students in the audio-tutorial method. Students with relatively low math aptitude (SAT Math scores of 625 or lower) and low math achievement, had higher predicted grades in the audio-tutorial method than did their counterparts in the standard method. Predicted grades of students having traits in the intermediate range of math aptitude and achievement did not differ significantly in the two methods. These research findings and the research design used in 1973 are described in detail in the accompanying paper, "A Trait-Treatment Interaction in a College Physics Course" (Appendix A). However, there are two points to be added to this report. The first is that math achievement was a more important trait in this research than was mathematical aptitude. That is, the math pretest scores accounted for a greater proportion of the variance in the grades than did the SAT math levels. The second point is that within the audio-tutorial method there were differences in achievement between the students assigned randomly and those assigned according to their preference. When math aptitude and math achievement were taken into account, audio-tutorial students who had been assigned randomly had significantly higher achievement than those assigned according to their preference. This result is suggested in Table 3 of the Trait-Treatment

Interaction paper* and was verified by means of a Johnson Neyman analysis of the interaction.

When the course was offered in the Spring semester 1974, we sought to determine whether the results of the previous year would be reproduced. To this end, we did not reveal these results to any of the course instructors other than the faculty member who had directed the course in 1973. We also tried to maintain fairly similar conditions in the course by again having each teacher teach a section in both methods, by having uniform quizzes, homework and exams, and so forth. However, changes beyond our control did occur. The major ones were as follows: (1) a different faculty member directed the course and lectured to the "standard" students than in the previous year; (2) the two interim exams were one and a half hours long instead of one hour; (3) the exams were constructed primarily by the course professor rather than by committees of instructors; (4) the lecturers utilized an "Instant Response System" which allowed the professor to learn student responses to questions by means of push-buttons at the students' chairs; (5) furthermore, the size of the course was decreased somewhat and the range of student abilities was thereby diminished because more students of high math aptitude and achievement had been encouraged to take the course during the previous term.

These changes are important in terms of comparing the research conducted in 1973 with that conducted in 1974. Taking each change in order:

(1) The change in professor decreased the degree of coordination between methods because the professor who gave the lectures in 1973 was one of the developers of the audio-tutorial material and the newly assigned professor had no experience with the audio-tutorial method of instruction. (This change may also have affected the difficulty level of the lectures. I believe that the professor in 1974 directed the lecture to students having lower ability levels than did the professor in 1973.)

* I am grateful to Lee J. Cronbach for pointing out this effect.

(2) The longer examination time period probably resulted in decreased time pressure in the exams in 1974 relative to 1973.

(3) Due to the fact that the exams were primarily constructed by one individual in 1974 and by committees in 1973, the exams in 1974 appear to have been more homogeneous and to have measured a narrower range of student learning than did those in 1973.

(4) The "instant response system" has been found by Professor R. Littauer[1] at Cornell to lead to increased student interest and to maintain higher levels of student attendance at lectures.

(5) The decreased range of student abilities affected both what was required of instructors in the classroom as well as the level of significance of statistical results.

These differences have been included here to help point out the difficulties inherent in attempting to reproduce a study of teaching, and to put in perspective the results of the 1974 research concerning achievement.

In 1974, the research concerning student achievement indicated major differences between the effects of random and preference assignment procedures:

(1) Among students who were randomly assigned to the methods there was a minor indication of an interaction similar to that found the previous year. That is, students who had higher math achievement did slightly better in the standard method than in the audio-tutorial method. Those with lower math achievement did slightly better in the audio-tutorial method than in the standard. However the difference in the slopes of regression lines and therefore the interaction that was obtained were not statistically significant. Furthermore, math aptitude as measured by SAT math scores was not highly related to grade in either method for randomly assigned students. (Please see Figure 1.)

1) R. Littauer, "Instructional Implications of a Low-Cost Electronic Student Response System," Educational Technology, XII (Oct., 1972), 69.

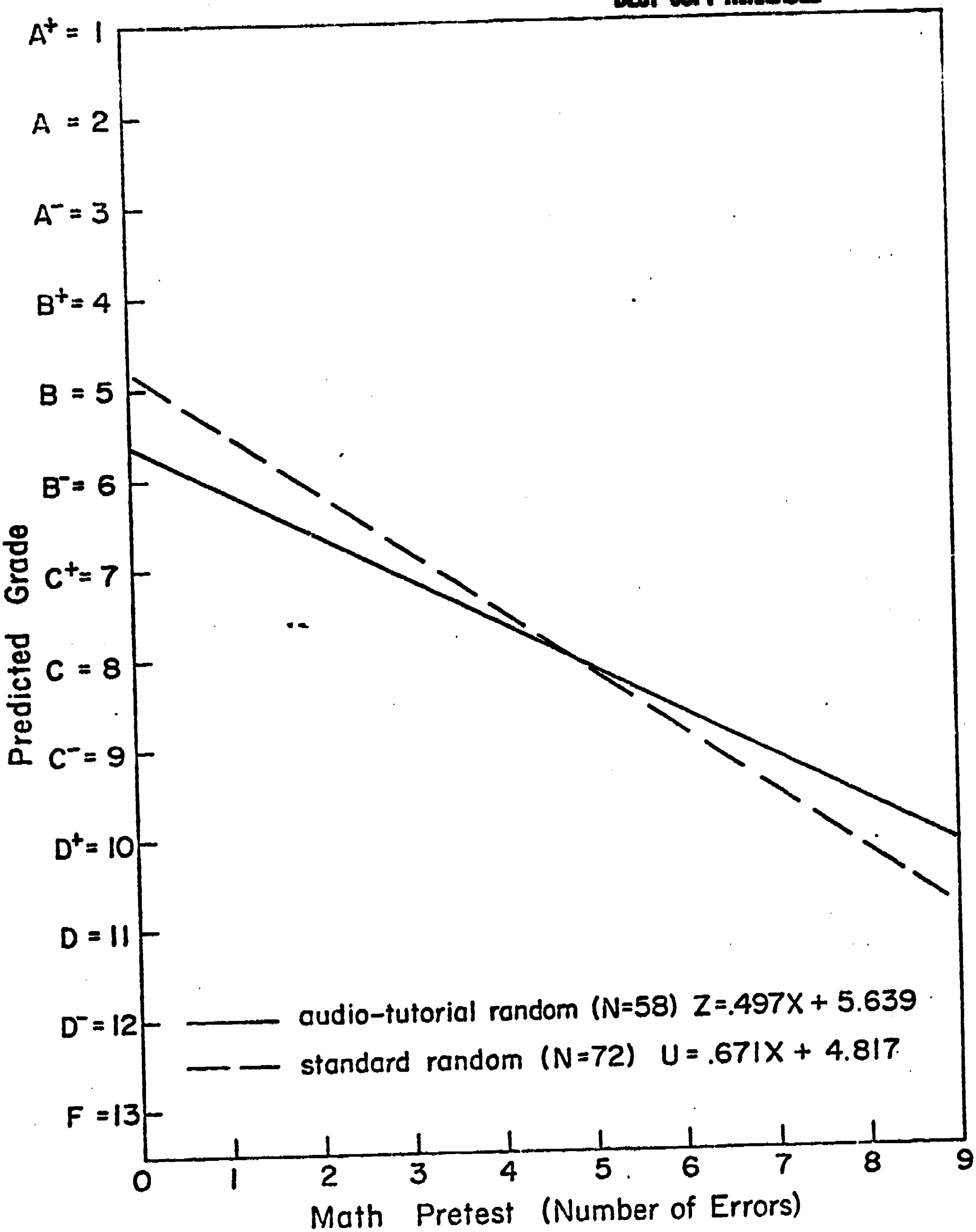

A+ = 1
A = 2
A- = 3
B+ = 4
B = 5
B- = 6
C+ = 7
C = 8
C- = 9
D+ = 10
D = 11
D- = 12
F = 13
Predicted Grade
audio-tutorial random (N=58) Z=.497X + 5.639
standard random (N=72) U = .671X + 4.817
0 1 2 3 4 5 6 7 8 9
Math Pretest (Number of Errors)

Figure 1

In combination with the 1973 results, these results indicate that there may be a tendency to a trait-treatment interaction for the standard and audio-tutorial methods and that this tendency may be reinforced to greater or lesser degrees under different course conditions and with different course instructors.

(2) Among students assigned by preference, those who selected the standard method clearly achieved higher final grades than did those who selected the audio-tutorial method.* Using the Johnson-Neyman technique to detect regions of significant differences between treatments, we compared the predicted grades for the audio-tutorial preference and the standard preference students. We obtained the following results:

a. For the range of student traits in our sample, the predicted grade was higher in every case for the standard preference group.

b. For a range of SAT Math values of about 630-670, the predicted grades were significantly higher in the standard preference group (level of significance = .05). (Note that these SAT math values are the actual scores received by the students, rather than levels as reported in 1973.)

These results are sketched in Figure 2. Statistics are given in Table 1. In Figure 2 the $c\eta$ line is the line indicating equal predicted grades in the two treatments. Above this line, predicted grades are higher in the standard preference group. Below this line, predicted grades are higher in the a.t. preference group. Due to the fact that no students had SATM scores which fell below the $c\eta$ line, the standard preference treatment had higher predicted grades than the a.t. preference treatment for all students in the population. In addition, for an elliptical region bounded by SATM scores of 630 to 670, predicted grades were significantly higher in the standard preference treatment than in the a.t. preference treatment (level of significance = .05). Thus the

* Note that the preference students who were assigned to a method contrary to their preference are not included in these and subsequent analyses. However such students were included in the 1973 analysis.

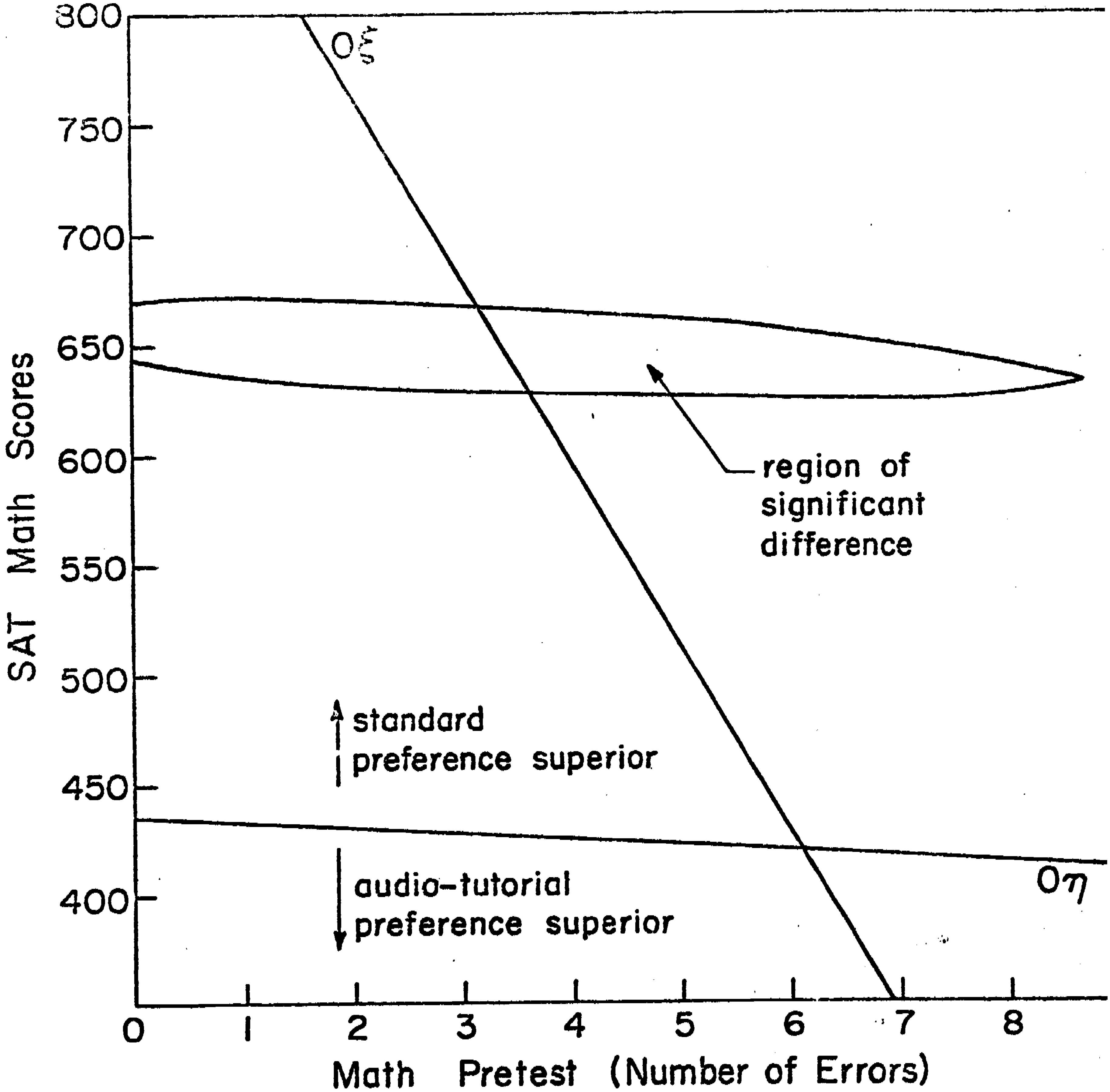

800
750
700
650
600
550
500
450
400
SAT Math Scores
0ξ
0η
region of significant difference
standard preference superior
audio-tutorial preference superior
0
1
2
3
4
5
6
7
8
Math Pretest (Number of Errors)

Figure 2

Table 1

Basic Statistics for Johnson-Neyman Analysis

	Treatments	
	Audio-tutorial Preference	Standard Preference
	$N_1 = 52$	$N_2 = 88$
Regression Equations	$z = 14.283+.527x-.012y$	$u = 16.201+.516x-.017y$
Means	$\bar{x}_1 = 2.50$ $\bar{y}_1 = 689.6$ $\bar{z} = 7.04$	$\bar{x}_2 = 2.68$ $\bar{y}_2 = 674.0$ $\bar{u} = 6.23$
Standard Deviations	$\sigma_{x'} = 1.67$ $\sigma_{y'} = 54.6$ $\sigma_z = 3.22$	$\sigma_{x''} = 1.71$ $\sigma_{y''} = 54.8$ $\sigma_u = 3.19$
Correlation Coefficients	$r_{xy'} = -.437$ $r_{xz} = +.365$ $r_{yz} = -.329$	$r_{xy''} = -.530$ $r_{xu} = .430$ $r_{yu} = -.436$

x = errors on math pretest

y = math SAT score

z = final grade in the audio-tutorial preference treatment

u = final grade in the standard preference treatment performance of the standard preference students surpassed that of the a.t. preference students. It is of interest to note that the a.t. preference students received grades in their other courses which were lower than the grades achieved by each of the three other groups of students, (standard preference, a.t. random, and standard random). Thus lower performance in P112 was not counterbalanced by higher performance in other classes.

Assuming that there were no systematic differences between students given the random or preference assignment procedures, the difference in course grades between a.t. preference and standard preference students and the lack of such differences within the randomly assigned groups require an explanation. It apparently is to be found in terms of the motivation of the students who selected the a.t. method. First of all, what were these students like in relation to those who chose the standard method?

From the initial questionnaire we learn:

Twelve of the fifty-two (23%) who chose a.t. and stayed with their choice had taken two or more years of high school physics, nine of the ninety (10%) who chose standard had also taken two or more years of high school physics. Fifty of the fifty-two (96%) who chose a.t. said that they would succeed well or very well in a science course without pressures from teachers or other students. Only fifty-nine of the ninety (66%) who chose standard so stated.

Twenty-eight of the fifty-two (54%) who chose a.t. thought P112 would be difficult, whereas sixty-three of the ninety (70%) who chose standard thought P112 would be difficult or very difficult.

None of the students who chose a.t. was registered in Math 191, whereas eight standard students were registered in this math course. Math 191, a course taught in an "exam tutorial", self-paced format, is usually a prerequisite to enrollment in P112. Apparently students who were behind in mathematics and who may have faltered in the non-traditional mode of instruction in M191 did not choose the audio-tutorial method in P112.

In regard to their previous experience in science and math courses, only eight of the fifty-two (15%) who chose a.t. thought recitations had been very helpful in the past. In contrast, thirty-nine of the ninety (43%) standard students thought recitations had been very helpful.

As far as math achievement and aptitude are concerned, those who chose a.t. did slightly better on the math pretest (2.5 errors vs. 2.7) and had higher average math SAT scores (690 vs. 674). Mean SAT math scores, math pretest scores, and final grades are given for each treatment group in Table 2.

As far as students' experiences in P112 are concerned, on a final questionnaire students reported the following:

Seven of the forty-five a.t. students responding (16%) said they stopped regularly attending the learning center by the fourth week of the course. None of the seventy-four responding students in the standard preference sections claimed to have stopped attending lecture that early in the semester.

A.t. preference students appear to have spent fewer hours in the learning center than a.t. random students. Moreover, twenty-four of forty-four a.t. preference students (55%) said that they did not regularly attend the learning center. The proportion in the a.t. random groups was fourteen out of forty-four (32%).

A.t. preference students said they spent about five hours per week outside class on P112, vs. 8.6 hours for standard preference students.

As far as achievement is concerned:

The standard preference students had a considerably higher average grade on the first prelim than did those who chose a.t. Other test grades did not differ significantly.

Finally, as mentioned earlier, the a.t. preference students had slightly lower grade point averages in their other courses than standard preference students.

Table 2

Means and Standard Deviations of Initial Traits and Final Grades*

Measure	Statistic	Within Assignment Procedure				All Sections	
		Random		Preference			
		a.t. (n=58)	standard (n=72)	a.t. (n=52)	standard (n=88)	a.t. (n=110)	standard (n-160)
SATM scores	mean	680.5	694.7	689.6	674.0	684.8	683.3
	s.d	56.1	59.0	54.6	54.8	55.1	57.3
Math Pretest (errors)	mean	2.88	2.90	2.50	2.68	2.70	2.73
	s.d	1.79	2.02	1.67	1.71	1.73	1.85
Final Grade (A+=1, F=13)	mean	7.07	6.76	7.04	6.23	7.05	6.47
	s.d.	3.07	3.12	3.22	3.19	3.12	3.15

* No differences in means were significant at the 5% level, two-tailed.

The picture which emerges of the "preference" students who chose the audio-tutorial method is that of a fairly confident group of students with a fairly good preparation in physics. Moreover they apparently had not found formal classroom instruction particularly useful in science and math courses in the past. They selected the a.t. method but in many cases made very little use of the available materials. Perhaps they put too much faith in their preparation in physics, as indicated by the low scores on the first test which mainly emphasized material treated in high school physics classes.

B. Student Attitudes

Student attitudes toward the audio-tutorial and standard methods used in P112 were evaluated by means of a questionnaire completed by students at their last recitation meeting of the semester. This procedure was followed in 1973 and in 1974. The questionnaires used in 1973 and in 1974 were similar in construction, involving evaluations of various aspects of the course and of the methods of instruction, as well as ratings of instructors.

In order to determine student attitudes toward the methods of instruction, two fairly global statements were included in each questionnaire. Students were asked to indicate the extent of their agreement with each statement by indicating either strongly disagree, disagree, neither agree nor disagree, agree, or strongly agree. The statements are as follows (with the appropriate method indicated in each case):

1) In general, I have been satisfied with the audio-tutorial (standard) method of instruction used in P112
2) I am glad I took the a.t. (standard) version of P112 rather than the standard (a.t.) version.

Student responses to these two statements are reported in Table 3 for both 1973 and 1974.

In general students tended to indicate greater agreement with the second statement than the first statement. That is, some students indicated that they were not satisfied with the method they received, but that they nonetheless were glad they took that method rather than the other method.

As far as assignment procedures are concerned, preference groups uniformly indicated greater agreement with both statements in both years than did random groups.

As far as methods of instruction are concerned, audio-tutorial students indicated greater agreement with the second statement in 1973 than in 1974. Standard students' opinions did not shift significantly from 1973 to 1974.

Table 3. Percentage of Student Responses to Questionnaire Statements.*

BEST COPY AVAILABLE

Statement 1. In general, I have been satisfied with the audio-tutorial (standard) method of instruction used in P112.

Year	Group	Total N	Percentage of Responses: Agree/Strongly Agree	Neutral	Disagree/Strongly Disagree
1973	ATR	51	41	16	43
	ATP	38	53	18	29
	STR	93	43	31	26
	STP	73	50	25	25
1974	ATR	44	30	27	43
	ATP	44	57	9	34
	STR	52	35	25	40
	STP	69	55	17	28

Statement 2. I am glad I took the a.t. (standard) version of P112 rather than the standard (a.t.) version.

Year	Group	Total N	Agree/Strongly Agree	Neutral	Disagree/Strongly Disagree
1973	ATR	52	48	27	25
	ATP	37	70	11	19
	STR	93	59	26	15
	STP	74	67	23	10
1974	ATR	44	39	20	41
	ATP	44	57	14	29
	STR	53	62	28	10
	STP	69	75	16	9

* Only those students who had initially preferred the method they were assigned to were included in the totals for preference groups. Students assigned to preference groups contrary to their preference tended to be very negative about the methods they received.

Standard and audio-tutorial students did not differ in the pattern of their responses to statement 1 either in 1973 or in 1974. Changes which occurred in responses to statement 1 consisted in less agreement with the statement among randomly assigned audio-tutorial and standard students in 1974 compared with 1973.

Responses to statement 2 indicated fairly similar responses among audio-tutorial and standard students in 1973. However there was decreased agreement with the statement among audio-tutorial students in 1974 relative to 1973 and slightly increased agreement among standard students in 1974 relative to 1973.

In summary:

1) Students assigned by preference expressed more positive attitudes toward the method they received than did students assigned by random procedures.

2) Audio-tutorial and standard students tended to be equally satisfied with the method they used (Statement 1).

3) In 1974, but not in 1973, standard students were more likely than audio-tutorial students to say that they were glad to have taken their method of instruction (Statement 2).

The greater disaffection with instructional method among audio-tutorial students may have been due to problems experienced in 1974 in coordinating the two methods of instruction, to increased problems with the availability of laboratory apparatus, to personnel changes, or to other causes.

In summary, therefore, we have learned that student achievement did not differ in 1974 among audio-tutorial random and standard random students. However, achievement was higher for standard preference students than for audio-tutorial preference students. In combination with the finding that in 1973 achievement was significantly higher for audio-tutorial random students than for audio-tutorial preference students, one has reason to question the advisability of allowing students to choose a method of instruction.

In regard to attitude, we have found that students assigned by preference were more satisfied with the method of instruction they received than were randomly assigned students. Furthermore, in 1974 audio-tutorial students were more likely to say they would have preferred the alternative method than were standard students.

III. Longitudinal Study of Student Achievement.

In order to determine whether there were longer-range effects of the differences in methods of instruction, we analyzed the progress of the engineering students in the 1973 sample (about 90% of the sample). We looked at the students in the a.t. and standard random sections (ATR and STR), and those in the a.t. and standard preference sections (ATP and STP). As in the research reported in Appendix A, we included only those students in the random sections who had been originally assigned to the recitation which they continued in, and only those preference students who actually obtained the method they preferred and and who did not transfer sections. Students who dropped P112 or transferred sections were excluded.* We sought to determine: 1) whether these groups of students enrolled in the same selection of engineering courses; 2) whether there were any negative effects of the audio-tutorial method when students returned to a lecture-recitation-lab format in their science courses; and 3) whether audio-tutorial and standard students had different attrition rates in later semesters.

A word is in order about the typical pattern of courses for sophomore engineering students at Cornell. Usually these students take one physics course (P213) in the first semester and two (P214 and P216) in the second. They also take two engineering courses each semester, usually from a selection of seventeen "basic studies" courses. We found however, that the approximately 240 engineering students in our sample who enrolled as sophomores enrolled in a total of 35 different engineering courses, including some upper division courses. On the other hand, only six engineering courses had substantial numbers of these students enrolled. These six courses were:

Computer Science 202 (Computers and Programming)
Engineering 9160 (Introductory Probability)
Engineering 4210 (Introduction to Electrical Systems)
Engineering 1021 (Mechanics of Solids)

* See p. 6 of Appendix A.

Engineering 6261 (Mechanical Properties of Materials)
Engineering 3631 (Introduction to Thermodynamics)

Percentages of the students from each research group in each of these courses are indicated in Table 4. It is apparent that there were no drastic differences in enrollment patterns from the four research groups.

Table 4. Enrollment in Engineering Courses (Percentage of Groups)

Course	ATR	ATP	STR	STP
Computers and Programming (202)	29	33	33	30
Introductory Probability (9160)	73	50	66	61
Introduction to Electrical Systems (4210)	63	70	58	58
Mechanics of Solids (1021)	51	47	57	64
Mechanical Properties of Materials (6261)	25	20	31	24
Introduction to Thermodynamics (3631)	29	33	22	28
Numbers in groups	51	30	86	74

Before turning to the achievement of the engineering students in engineering and physics courses, four points should be noted:

1) The STR group included 96 engineering students in Spring 1973, and the ATR group included 57. This standard random group had somewhat higher SAT math scores and somewhat lower math achievement scores than the a.t. random group. As attrition occurred during the next two semesters, these relationships were maintained.

2) The ATP group included 37 engineering students in Spring 1973 and the STP group included 83. This standard group had lower SAT math and math achievement scores than the a.t. preference group. As attrition occurred, the math achievement and the SAT math differences continued through Spring 1974 although the differences became attenuated.

3) The standard random engineering students had received significantly lower average grades in P112 than did those in the a.t. method. The audio-tutorial preference engineering students had received slightly lower grades in P112 than did those in the standard preference group and significantly lower grades than did those in the a.t. random group.

4) Finally, one must caution that it is somewhat risky to compare performance of the groups of students in the engineering courses because assignment to these courses was not random.

The only clear pattern which emerged from an analysis of student grades in these six engineering courses is that students from the standard random group had the lowest average grades in five of the six courses (that is, all but Engineering 1021 -- Mechanics of Solids).

In regard to physics courses, there was no apparent difference in the (small) numbers of engineering students from the four groups who enrolled in the honors version of these courses. In regard to the physics courses taken by most engineering students we obtained the following results:

Random sections: In P213 (Electricity and Magnetism), students in the standard random group received lower average scores than those in the a.t. random group. Nine students from the standard random group did not re-enroll in the subsequent semester. In P214 (Optics, Waves, and Particles) and P216 (Laboratory), there were no large differences in average grades.

Preference sections: In P213 there was no difference in average grades; in P214 the standard preference students had substantially higher grades than the a.t. preference students. In P216 the standard preference students had somewhat higher grades than the a.t. preference students.

Further, we analyzed the overall performance of the four groups of engineering students in their sophomore year. As the measure of overall achievement, we calculated the product of each student's grade point average times the total number of credits he took that semester. Comparing values of these products for the Fall semester 1973 and the Spring semester 1974, we obtained the following results:

During the first semester of the sophomore year, students from the a.t. random sections had somewhat higher achievement than standard random students. However after the attrition of nine of the standard random students and only one of the a.t. random students, there was little difference in the achievement of students from the two groups. On the other hand, during both semesters the standard preference students had higher achievement than the a.t. preference students.

Summarizing the course achievement results, it is clear that the randomly assigned audio-tutorial students performed at least as well as (and generally better than) the randomly assigned standard students in their subsequent courses. However, among students assigned by preference, the standard students performed somewhat better in some courses than did their audio-tutorial counterparts. That is to say, the relations among the P112 grades of these groups of students were reproduced to some degree in their later courses. Audio-tutorial random students out-performed standard random students; standard preference students out-performed a.t. preference students. We believe therefore that the audio-tutorial students as a whole were not adversely affected when they returned to lecture-recitation-lab format courses.

Finally, we will consider attrition rates of engineering students in each of these four groups. Please refer to Table 5 for the number of students in each of these four groups during each of the three semesters.

Table 5

Number of Students in Groups

Semester	ATR	ATP	STR	STP
Spring '73	57	37	96	83
Fall '73	51	30	86	74
Spring '74	50	27	77	72
attrition	7/57	10/37	19/96	11/83
% attrition	12%	27%	20%	13%

Thus the rate of attrition among a.t. random students was lower than among standard random students; a higher rate of attrition occurred among a.t. preference students than among standard preference students. Therefore rates of attrition were higher for those groups which had lower achievement in P112 (ATP, STR). Neither method was superior in regard to rates of attrition.

Summarizing the longitudinal study:

1) The audio-tutorial and standard students enrolled in the same selection of engineering courses;

2) There were no apparent differences in achievement favoring the standard students in relation to the audio-tutorial students;

3) Audio-tutorial and standard students had approximately equal rates of attrition.

IV. Observations of Instruction

This section is composed of two parts. The first deals with observations of learning center instruction. The second deals with observations of recitations and of lectures. Observations were conducted to delineate the characteristics of the two methods of instruction and thereby to make concrete the referent of this study and to increase its generalizability to other situations.

A. Learning Center Observations

The purpose of the observation of students and of tutors in the audio-tutorial learning center was to delineate the proportion of time students and tutors spent in the many kinds of activities available in the learning center. To this end we developed a fairly simple observation instrument and trained an observer in its use. (Please see Figure 3).

The instrument was used as follows:

During a period of 62 hours (distributed throughout the semester and representing each time period and day of week approximately equally), the observer made notations of the activities of the students and of the tutor. That is to say, every five minutes she noted how many students were engaged in each activity in Part 1 of the instrument, and which activity in Part 2 the tutor was engaged in.

The results were analyzed as follows:

1) In terms of the total number and percentage of students engaged in each activity under Part 1, both for the whole semester and for each of seven time periods during the semester;

2) In terms of the percentage of time spent by the tutors in each activity under Part 2;

3) In terms of the differences in time spent by the various tutors in each activity under Part 2.

4) We also used the count of students present in the learning center during the observations to estimate the total number of hours students spent in the center. This number was then compared with that obtained from tutor reports and from students' own accounts.

Figure 3

Learning Center Observations

(1) Student work

1. discussion of course materials
2. performing or looking at a demonstration
3. filling out a unit evaluation sheet
4. viewing a film loop
5. doing homework
6. doing lab work
7. doing miscellaneous, non-course work, etc.
8. taking a pre-test
9. doing a lab report
10. waiting to talk to tutor
11. reading text

Study Guide Work

12. reading objectives
13. working on portions other than problems
14. just listening to tape
15. looking at slide and listening to tape
16. looking at suggestion slide
17. working out problem solution
18. looking at solution slide
19. checking or correcting one's own solution
20. copying solution without attempting problem first
21. taking or checking a self-test
22. referring to checklist

(2) Tutor-Student work, or tutor alone

1. tutor alone at desk or elsewhere, not in contact with students or equipment
2. discussion of course material
3. helping student with demonstration
4. tutor fixing equipment
5. discussion of homework
6. helping with lab activity
7. discussing miscellaneous, non-course topics
8. discussing a quiz
9. helping with lab report
10. walking around learning center, seeking questions
11. out of room
12. on phone
13. listening to tape recording
14. at blackboard

I. Student Activities.

In all, 4794 notations were made to describe individual student's activities. Each notation referred to the activity of a single student. The major activities comprising 82% of the total were as follows: (see Figure 4)

1) 26% = lab work (#6)
2) 20% = working out problem solutions in study guide (#17)
3) 15% = listening to tape recordings (#14)
4) 10% = doing miscellaneous, non-course work, entering or leaving the room, etc. (#7)
5) 6% = discussion of course materials (generally with tutor)(#1)
6) 5% = looking at a slide giving a solution to a problem in the study guide (#18).

The observations of student activities were also analyzed for each of seven time periods during the semester. These periods were as follows:

1) Feb. 4-10 = first week
2) Feb. 11-19 = period immediately prior to first prelim
3) Feb. 20 - Mar. 8
4) Mar. 9-24
5) Mar. 25 - Apr. 2 = period immediately prior to second prelim
6) Apr. 3-23
7) Apr. 24 - May 16 = end of semester

As Figure 5 indicates, there was a greater amount of student time spent on activities in the study guide relative to that spent on lab activities prior to each prelim and the final exam. The figures also indicate a decrease in the average number of students present in the learning center as the semester progressed. These findings are in accord with informal observations of the learning center.

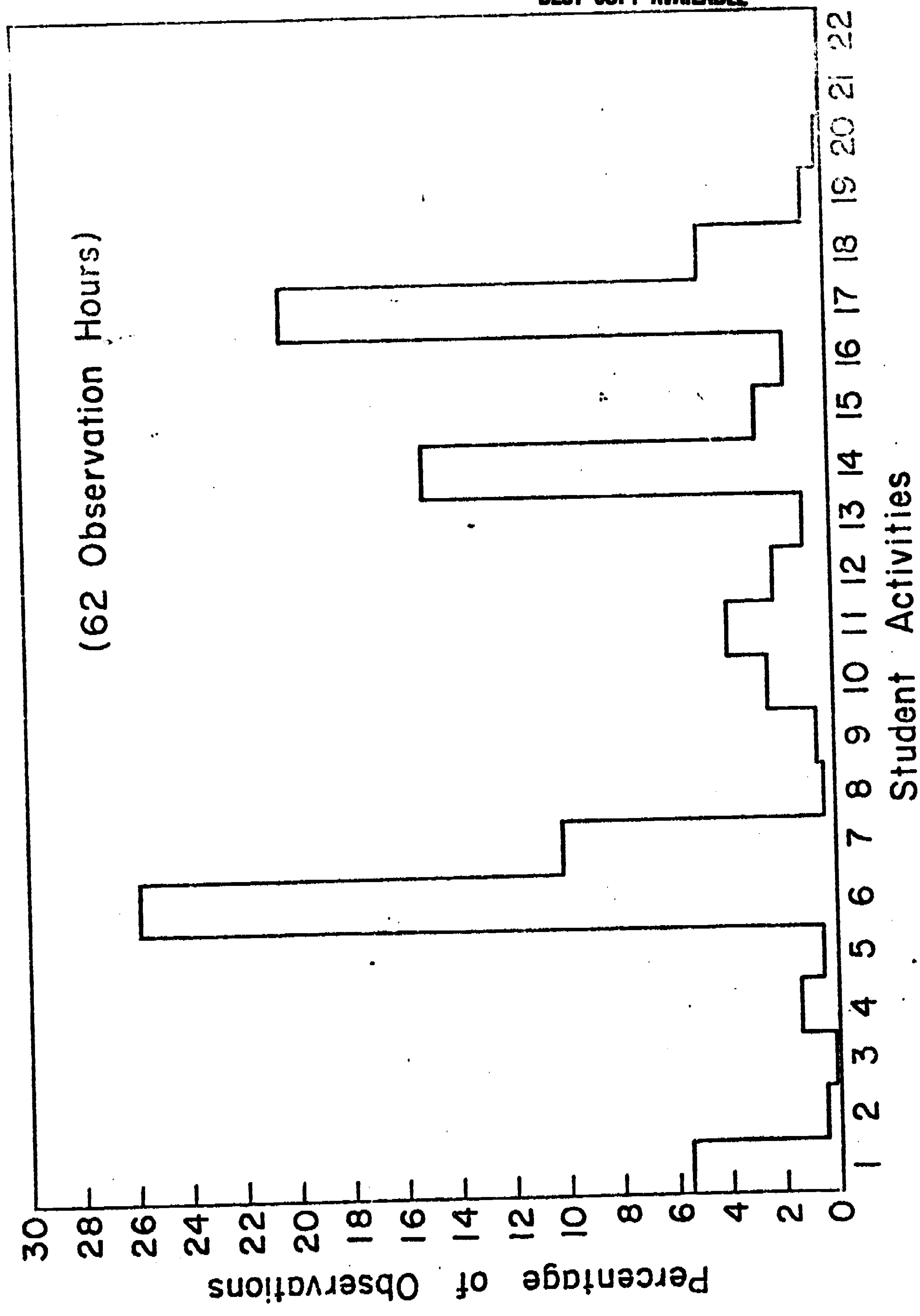

(62 Observation Hours)
Percentage of Observations
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Student Activities

Figure 4

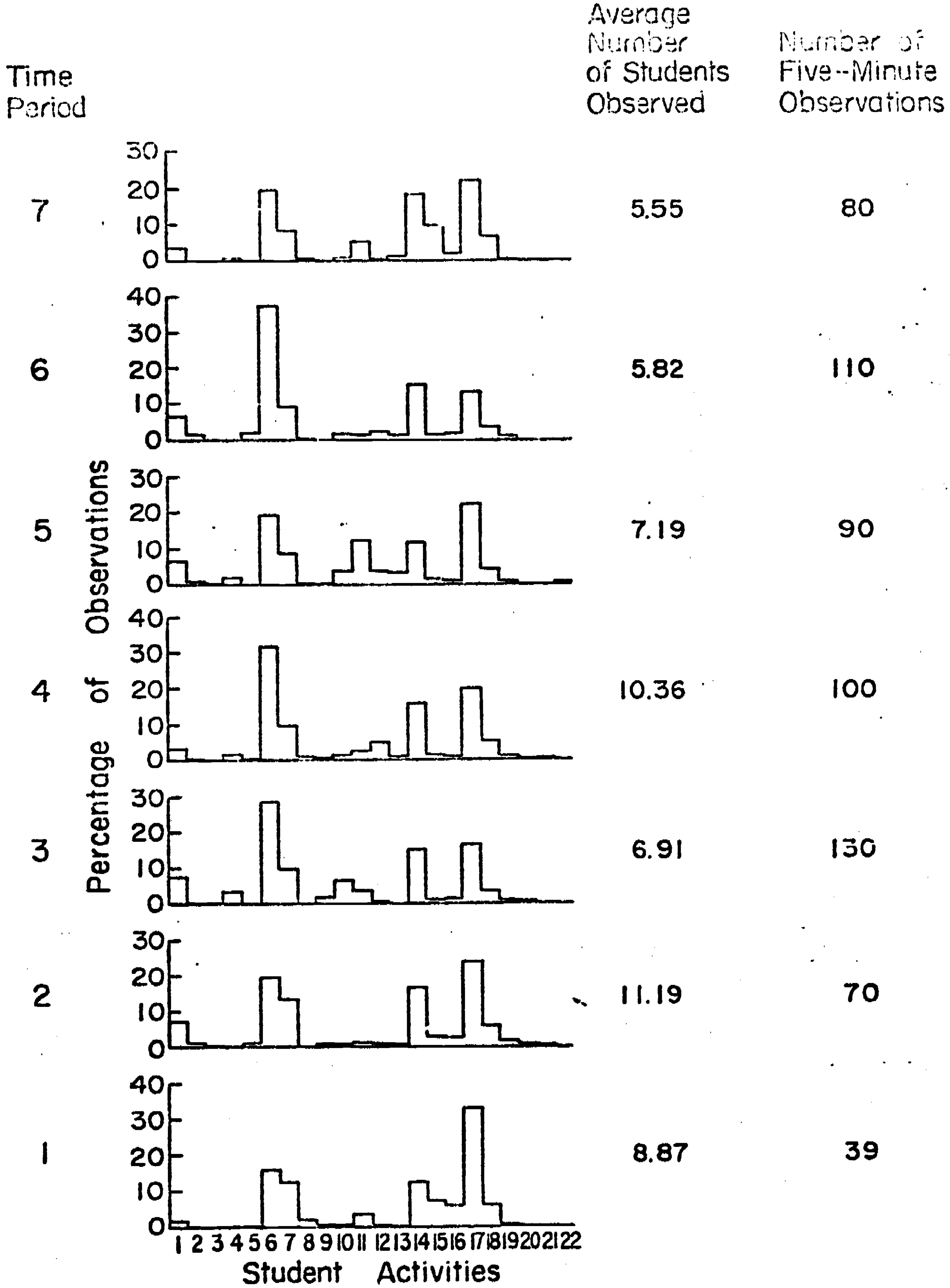

Time Period
Average Number of Students Observed
Number of Five-Minute Observations
7
5.55
80
6
5.82
110
5
7.19
90
4
10.36
100
3
6.91
130
2
11.19
70
1
8.87
39
Percentage of Observations
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Student Activities

Figure 5

2. Tutor Activities.

Eleven persons served as learning center tutors. These included seven instructors (two faculty members and five physics graduate students) who taught audio-tutorial and standard recitation sections, three graduate students who taught standard lab and graded audio-tutorial students' lab reports, and a faculty member who had helped to develop the audio-tutorial materials (Dr. Ott.)

One tutor was scheduled in the learning center at each hour of its operation. However, occasionally a second tutor would also be present. Thus there were 658 tutor activities noted during 619 observation periods.

The major activities comprising 87% of the total were as follows: (See Figure 6):1) 33% = tutor alone at desk or elsewhere, not in contact with students or equipment (#1)

2) 31% = tutor discussing course material [with a student](#2)

3) 14% = tutor walking around learning center, seeking questions (#10)

4) 9% = tutor helping students with lab activity (#6)

3. Differences Among Teachers.

The various tutors behaved quite differently in the learning center. Four tutors spent about half the observed time completely out of contact with students (#1). Two spent about half the time discussing course material (#2). Two spent about 20% of the observed time helping students with lab (#9) and three spent about 30% or more of the time walking around the learning center seeking questions (#10).

BEST COPY AVAILABLE

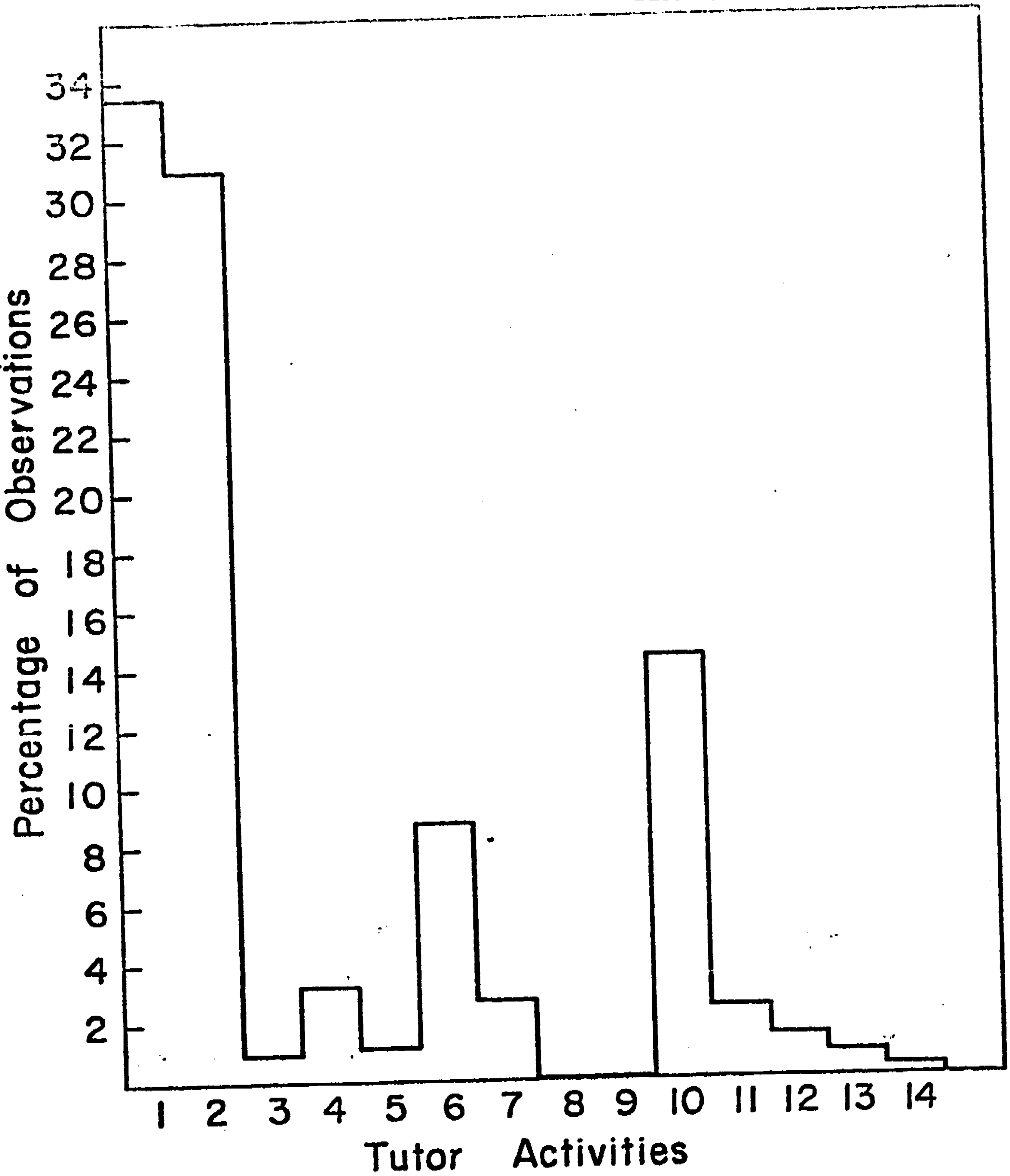


Figure 6

Minimum and maximum percentages of total observed time for individual tutors in each activity in Part 2 of the instrument were as follows:

	Tutor Activities	Minimum	Maximum
1)	tutor alone at desk or elsewhere, not in contact with students or equipment	10%	55%
2)	discussion of course material	20	51
3)	helping student with demonstration	0	2
4)	tutor fixing equipment	0	22
5)	discussion of homework	0	4
6)	helping with lab activity	4	24
7)	discussing miscellaneous, non-course topics	0	13
8)	discussing a quiz	0	0
9)	helping with lab report	0	0
10)	walking around learning center, seeking questions	2	36
11)	out of room	0	8
12)	on phone	0	4
13)	listening to tape recording	0	7 (only one tutor re-
14)	at blackboard	0	3 presented)

4. Number of Student Hours in Learning Center.

An estimate was obtained of the average number of students present in the learning center during the observation periods. That number is 4794 students/619 five-minute observation periods, or 7.74 students/observation period.

On the basis of counts made by the tutors of the number of students present at each hour in the learning center, the average number of students present was 6.78 -- in good agreement with that obtained by the observer.

These two numbers contrast with the total number of hours reported by students on "time cards" which they were asked to complete each time they worked in the learning center. These totaled 2725 hours for 140 students. Averaged over the approximately 640 hours in which the learning center was open, one obtains only 4.3 students/hour on the basis of "time card" reports. Thus students apparently failed to record about one third of the hours they spent in the learning center.

From the student time cards, one calculates that the average student spent about twenty hours in the learning center during the semester or 1.4 hours/week. On the basis of tutor and observer counts of students, one can assume that this number is closer to thirty hours per semester or 2.1 hours/week.

B. Observations of Lecture and Recitation Instruction

We were interested in observing both cognitive and behavioral aspects of physics lecture and recitation instruction, with our primary concern being the cognitive realm. However, a thorough review of the literature did not reveal any observation instrument suitable for observing cognitive behavior in a college physics classroom. The author therefore developed such an instrument, patterning it, to some degree, on the Wright-Proctor System[2] which was designed for use in the mathematics classroom.

The instrument is intended for use in a physics classroom. It has three components -- a list of "content" categories (please see Appendix B) which refer to the function of the statements being made by the teacher or by the students, a list of "behavior" categories (Appendix C) which refer to the non-verbal behavior of the teacher as well as verbal interactions between teacher and students, and a listing of the physics topics being discussed. (These topics were noted but not analyzed.) The two types of categories, content and behavior, correspond to an attempt to consider both cognitive and behavioral aspects of physics instruction.

Furthermore, audio tape recordings were made of most of the classes which were observed. Tapescripts of a selection of these recordings will be analyzed as a further probe into the cognitive aspects of physics instruction.*

Use of the observation instrument requires some knowledge of the terms and forms of argument used in physics. Therefore a background in physics was a qualification for observers. Two persons with at least two years of college physics instruction were selected to be observers. They were then trained to use the observation instrument by means of audio and video tapes of physics lectures and recitations. A brief listing of the instructions given to the observers is contained in Appendix D. The type of form used for recording the observations in included in Appendix E.

2. Muriel J. Wright and Virginia Proctor, "Wright-Proctor System" in Mirrors for Behavior II: An Anthology of Observation Instruments (Philadelphia Research for Better Schools, Inc., Spring 1970) p. 26-3.

* Please see Appendix F for a report concerning the relation between cognitive aspects of instruction and student achievement.

A systematic design of observation was devised. Each of the observers was assigned to observe selected lecture and recitation classes in P112. All observed lectures were at the second of the two scheduled lecture hours (i.e., 12:20 p.m.). In addition, the observers each attended fourteen different recitation sections at various class hours. Seven teachers taught these fourteen sections. Each teacher taught one standard and one audio-tutorial recitation section. Standard recitation sections were attended by students in the standard method of instruction. These met for fifty minutes twice a week for fourteen weeks. Seven standard sections were observed twelve times during the semester, six times by each observer. The seven audio-tutorial sections were attended by students in the audio-tutorial method of instruction. These sections met for fifty minutes once per week. Each of these sections was observed six times, three times by each observer. Each section was observed for about 43% of the class meetings.

The main functions of the observation instrument were to 1) specify the nature of lecture and recitation instruction in this course, 2) determine the extent and kinds of variation among recitation instructors and between standard and audio-tutorial recitations, and 3) provide a basis for determining the interrelations among student achievement, student evaluations of teachers, and teacher behaviors. Each of these functions will now be discussed.

1. Lecture Observations and Recitation Observations

Lectures were observed by at least one observer and frequently by both observers. All twenty-six lectures were observed. In order to compare observers' categorizations of the lectures, observations of a set of thirteen lectures from one observer were compared with observations of thirteen other lectures from the second observer. The quantity analyzed was the total number of times each content category and each behavior category were noted for each lecture observed. For each content category and each behavior category, the two observers' totals for each of thirteen lectures were compared using the methods of analysis of variance.

It should be noted, however, that content categories 8, 9, and 10 were combined in the analysis due to the similarity in the cognitive level of these categories and to the small number of observations in each category. (Category 8 refers to organizing or structuring the material, category 9 to specializing to a particular case, and category 10 to generalizing to a more inclusive treatment from a particular case. Note that 9 and 10 just indicate how often these categories were used rather than the total amount of time spent in these categories. For example, the continuation of a generalization would be noted in other categories, depending on the further content of the discussion.)

There were no significant differences between the two observers on any of the content categories. There was, however, one behavior category for which differences between observers were significant at .05. That category was category E which refers to erasing the blackboard, moving a view-graph, or removing a transparency from the view-graph. Thus, excepting category E, the observers agreed in terms of lecture observations.

The mean values of the lecture observations are given in Table 6. Recall that two observations were made every minute. Thus a mean value of 11.35 for category F indicates that, on the average, 11.35/2 or 5.68 minutes of each lecture were included in category F. (Category F refers to the teacher performing a demonstration or showing a movie.) As indicated in Table 6, the predominant content category in lecture was category 3, the description of a physical system or process or demonstration or the use of a diagram or graph to illustrate a process or kind of motion. That is, lecture emphasized demonstrations.

Second in usage was category 7, deriving or explaining an equation or formula with explicit reference to its physical meaning or implications. The only other content category which received an average of more than six notations per lecture was the combined category formed from 8, 9, and 10. These deal with organizing or structuring the material, generalizing, or specializing.

BEST COPY AVAILABLE

Table 6

Mean Values of Lecture Observations

Content Categories		Behavior Categories	
0	5.15	0	28.19
1	4.04	A	.12
2	5.81	B	26.88
3	35.04	C	.04
4	5.58	D	13.35
5	5.81	E	2.27*
6	1.96	F	11.35
7	18.42	G	2.00
8, 9, 10	11.73	H	.38
11	1.65	I	5.58
		J	5.04

* observer disagreement significant at .05

Table 7

Mean Values of Recitation Observations

Content Categories	all sections	a.t.	standard	Behavior Categories	all sections
0	*	11.98	6.51	0	37.50
1	7.28			A	27.39
2	4.82			B	
3	25.63			C	12.79
4	8.44			D	
5	4.41			E	1.73
6	7.44			F	.63
7	24.69			G	17.33
8, 9, 10	**	5.71	9.15	H	2.09
11	.14			I	
				J	

* difference between methods significant at .05

** difference between methods significant at .01

The predominant behavior category was category 0 which simply indicates that none of the other behavior categories was appropriate. For lecture, category 0 primarily includes periods of time in which the instructor is simply speaking to the class rather than doing a demonstration, writing on the blackboard, etc.

The second most common behavior category, nearly equal in frequency to the first, is category B, writing an equation or words on the viewgraph (overhead projector), or showing a prepared transparency of an equation or of words on the view-graph.

Category D, drawing a diagram on the viewgraph or showing a prepared diagram on the viewgraph, was third in frequency. Category F, performing a demonstration or showing a movie, was fourth.

Categories I and J both referred to the use of the Instant Response System. This is a push-button system by means of which students' responses to questions are recorded. The system was available only in the lecture hall used on Mondays. Therefore the observations listed under I and J represent behaviors observed at only half of the lecture classes. Thus although an average time of about five minutes per lecture is indicated in Table 6 for categories I plus J, there was no time spent on these categories on Mondays and about ten minutes spent in these categories in the Friday lectures.

Recitation observations required a more complicated analysis than did lecture observations. There were three primary factors in the analysis: 1) method (audio-tutorial or standard), 2) observer, and 3) teacher. That is, there were two methods, two observers, and seven teachers. The detailed results of the analyses of variance will be described in section 2. However, the averages across teachers, methods, and observers will be cited here for comparison with the lecture observations. These are listed in Table 7. These averages represent observations of one hundred twenty-six entire recitation sections. The averages were formed as follows: the average for eighty-four standard recitation observations and the average for forty-two audio-tutorial recitation observations were then averaged to form the average for all recitations. Thus standard and audio-tutorial recitations received equal weight in the final averages. It should be noted that individual teachers differ significantly from the average on a number of these categories.

As noted in Table 7, there were significant method differences for content categories 0 and the composite category 8, 9, and 10. Thus for these categories the averages for each method are listed rather than the overall average. The explanation for these differences is apparently that audio-tutorial recitations met only half as often as standard recitations but both types of recitations had an equal number of quizzes. Thus a higher proportion of class time in audio-tutorial sections was devoted to quizzes (which were categorized as 0). Apparently the time pressures which resulted from fewer class hours caused the teachers to use the higher order cognitive areas included in categories 8, 9, and 10 less in a.t. sections than in the standard sections. No averages are given for categories B, D, I, and J in Table 7 because view-graphs and the Instant Response System were not available to recitation instructors.

As indicated in Table 7, the predominant content category used in recitation was category 3. Category 7 was next in usage, nearly equaling category 3. These are followed by category 0 in audio-tutorial recitations and by 8, 9, and 10 in standard recitations.

In terms of behavior categories, the most common ones used in recitations were 0 (none of the others), A (writing an equation or words on the blackboard), G (questioning or discussion with one student), and C (drawing a diagram on the blackboard).

Lecture and recitations, therefore, tended to be similar entities in some respects. The same content categories predominated in lecture and in standard recitations, although lecture emphasized descriptions of systems (category 3) rather than derivations or explanations of formulas with reference to physical meanings or implications (category 7). The "average" recitation had approximately equal time devoted to these two areas.

As far as behavior categories are concerned, category 0 predominates in both lecture and recitation. Category A is analogous to category B, One notes that these categories are next in usage in the two forms of instruction. These refer to writing an equation or words on the blackboard (A) or on a view-graph (B). Each form of instruction included about the same amount of drawing of diagrams on the blackboard (C) or on the view-graph (D). However, lecture included a far greater emphasis on demonstrations (F) than did recitations. Conversely, recitations tended to emphasize questioning or discussion with individual students (G) and this category was infrequently noted in lecture. Note that this is to be expected considering the purpose of the two forms of instruction.

Thus lecture and recitations on the average had many features in common. However, as will be detailed in the next section, the average values given for recitations in Table 7 do not represent the diversity of teaching styles which were observed in the recitations.

2. Observation of Recitation Instruction: Method, Observer, and Teacher Effects

As suggested earlier, the analysis of recitation observations was far more complex than the analysis of lecture observations. The research design for recitation observations had three factors -- method, observer, and teacher. There were two methods, two observers, and seven teachers. Each teacher was observed eighteen times, six times by each observer in the teacher's standard recitation section, and three times by each observer in the teacher's audio-tutorial recitation section. Thus there was a total of 126 observation periods, equally divided among teachers and among observers. However, audio-tutorial classes were represented in forty-two (1/3) of the observations, and standard classes were represented in eighty-four (2/3) of the observations.

In analyzing each separate content and behavior category we employed analysis of variance techniques. The quantity analyzed in each case was the total for a category for an entire recitation period.

A sample analysis of variance table, indicating the sources of variation and degrees of freedom, is presented here to indicate the factors and the two- and three-factor interactions which were analyzed.

Source	df
Method	1
Observer	1
Teacher	6
Method by Observer	1
Method by Teacher	6
Observer by Teacher	6
Method by Observer by Teacher	6
Error	98
Total	125

For categories in which method by teacher interactions were found to be significant, we revised the analysis of variance so as to consider differences between recitation sections. That is, teacher and method factors were grouped into a factor called "section". The analysis of variance table for this alternate analysis is indicated below:

Source	df
Observer	1
Section	13
Observer by Section	13
Error	98
Total	125

A significant method by teacher interaction and thus a significant section effect means that individual teachers differed in the ways they taught their audio-tutorial sections relative to the ways they taught their standard sections.

The remainder of this section will treat the content categories and then the behavior categories, looking at method differences, observer differences, and teacher differences. Interactions will be mentioned if significant. The level of significance was set at .05.

a) Content Categories

Category 11 is not analyzed because events were too rare to be stable.

Interactions

For category 1, there was a method by teacher interaction significant at .01. Thus this category was analyzed by section rather than by method and by teacher. These results are discussed under section differences.

No other interactions were significant.

Method Differences

The only method differences were as follows:

For category 0 there was a difference between methods, significant at .05. Category 0 was noted more often for audio-tutorial than for standard sections. As mentioned in the section comparing lecture and recitation, this difference is attributable to the fact that audio-tutorial sections spent proportionally more time taking quizzes than did standard sections.

For the combined category 8, 9, 10 there was a difference between methods significant at .01. Category 8, 9, 10 was more common in standard than in audio-tutorial recitations. Apparently teachers reacted to the fact that audio-tutorial sections had half as much recitation time by using these higher order cognitive categories less often in audio-tutorial sections than in their standard sections.

Teachers appear to have treated audio-tutorial and standard sections in fairly similar ways. However, they deemphasized higher order cognitive categories 8, 9, and 10 in audio-tutorial sections.

Observer Differences

Observer differences were noted for category 1 (significant at .05), for category 3 (significant at .05), and for category 4 (significant at .01). These differences indicate a need to clarify distinctions among categories. Fortunately, however, these observer differences did not significantly interact with method or teacher effects.

Teacher Differences (Please refer to Table 8.)

For purposes of comparison teachers will be referred to as I, II, III, IV, V, VI, and VII. Significant differences were found in two content categories.

Teacher differences significant at .01 were found for category 4 which relates to stating or rephrasing a problem. Teachers I and IV had significantly fewer observations in this category than teachers III, V, and VII. (The Neyman Keuls analysis was used to test differences.) Teacher differences significant at .01 were found for category 7 which is concerned with describing or explaining an equation or formula with explicit reference to physical meaning or implications. Teacher I was significantly higher on this category than teachers II, III, V, or VII.

These results for categories 4 and 7 indicate that some teachers (I and perhaps IV) emphasized the statement of problems more than others did (III, V, VII, and perhaps II). Further they indicate that those teachers who emphasized problem statements, de-emphasized problem solutions relative to the other teachers. Finally, teacher VI did not emphasize either category relative to other teachers.

One should note that all teachers had more observations in category 7 than in category 4. However, the ratio between the observations in these categories varied from

$32.44/2.94 \approx 11.0$ for teacher I to

$22.67/10.67 \approx 2.1$ for teacher V.

Table 8

Mean Values of Content Categories for Teachers

Content Categories / Teacher	I	II	III	IV	V	VI	VII
0	3.83	9.56	6.78	7.83	10.83	10.67	8.83
1** standard	4.25	6.15	8.17	4.75	4.75	7.40	6.25
1** a.t.	12.83	13.50	4.33	4.83	6.83	12.33	5.50
2	4.62	4.28	3.33	4.33	5.72	6.11	4.50
3	26.28	25.94	29.61	27.44	26.28	19.61	28.06
4*	2.94	9.61	10.33	4.89	10.67	8.56	11.94
5	4.22	9.00	4.44	4.94	3.89	3.50	3.67
6	4.28	5.94	6.33	9.50	6.61	11.28	7.50
7*	32.44	19.28	20.22	28.17	22.67	26.94	23.55
8,9,10	9.78	10.78	7.50	6.56	7.39	6.94	7.11

* Teacher difference significant at .01

** section difference significant at .01

Section Differences

Category 1, which involves statements concerning course and classroom administration, was analyzed by recitation section due to a significant method by teacher interaction. Differences between sections were significant at .01. All three of the sections which were high on category 1 were audio-tutorial sections. These were taught by teachers I, II, and VI.

Sections significantly lower than these on category 1 were the standard sections taught by I, IV, and V, and the audio-tutorial sections taught by III, IV, and VII. In addition, the standard section taught by II was significantly lower on category 1 than teacher II's audio-tutorial section.

Apparently a number of teachers used category 1 proportionally more of the time in their audio-tutorial than in their standard classes. However, other teachers apparently de-emphasized this area in their audio-tutorial classes.

b) Behavior Categories

Categories B, D, I, and J did not pertain to recitation instruction.

Interactions

There was a method by teacher interaction for category A which was significant at .10. For this reason category A was analyzed by sections.

There was an observer by teacher interaction for category F which was significant at .01, and an observer by teacher interaction for category H which was significant at .05. The observer and teacher main effects for these categories will therefore be considered in the light of these interactions.

Method Differences

There were no behavior categories for which method differences were statistically significant (at .05).

Observer Differences

For both category F and for category H there were observer differences significant at .01. (F is concerned with demonstration, H with discussions between the teacher and two or more students.) Observer 2 used category F more often than did observer 1. Observer 1 used category H more often than did observer 2. However, these differences were not consistent across teachers as indicated by the presence of observer by teacher interactions.

Teacher Differences (Please refer to Table 9.)

Teachers differed on category 0 (none of the other behaviors), significant at .05. The teachers highest and lowest in this category were teacher VI and teacher VII, respectively.

Category A (writing an equation or words on blackboard) is described under Section Differences due to the method by teacher interaction.

Teachers differed on category C (drawing a diagram on blackboard), significant at .05. Teacher VI was significantly lower on this category than was teacher II.

Teachers differed on category G (questioning or discussion with one student) significant at .01. Teacher I used this category significantly less than all other teachers. Furthermore, teachers IV, II, and VI used this category less than did teachers V and VII.

Category F and category H each reflected significant differences between teachers but were complicated by observer by teacher interactions. However, for category F observers agreed that teacher VII was significantly higher than all the others. There was disagreement on the relative amount that the other teachers used category F, but all were substantially lower than teacher VII.

For category H there is a substantial disagreement among observers about individual teachers which makes teacher differences unreliable.

Table 9

Mean Values of Behavior Categories for Teachers

Behavior Categories / Teacher	I	II	III	IV	V	VI	VII
0	39.17	42.33	31.28	38.22	32.33	44.56	31.11
A*** standard	35.25	26.25	24.67	35.08	22.25	29.33	28.67
A*** a.t.	28.50	19.50	30.00	24.00	24.83	31.00	24.00
C*	15.00	16.22	11.33	12.78	14.94	9.67	11.33
E	1.61	1.50	2.50	1.72	1.56	2.17	1.44
F**	.33	.78	.22	.06	.22	.28	2.83
G**	5.72	14.50	20.56	13.56	23.67	14.72	25.28
H	(.72)	(1.83)	(3.39)	(.89)	(3.61)	(1.39)	(2.33)

* teacher difference significant at .05
** teacher difference significant at .01
*** section difference significant at .01

Summarizing significant teacher differences:

1) for category 0, VI used this category more than VII did;
2) for category C, VI used this category less than II;
3) for category F, VII used this category more than all other teachers
4) for category G, I used this category less than all other teachers and in addition, IV, II, and VI used the category less than did teachers II and VII.

Section Differences

Category A was analyzed by section due to the method by teacher interaction. The difference between sections was significant at .01. The difference found was that teacher II used this category significantly less for his audio-tutorial section than teachers IV and I did for their standard sections. Other sections did not differ significantly. There was no major pattern of differential usage in audio-tutorial vs. standard sections.

To complete this section on recitation observations, we wish to stress the differences in teacher behaviors which make it difficult to specify the nature of recitation instruction. To do this we will indicate the ranges of mean values on each of the content and behavior categories for the teachers. These are given in Table 10 in minutes per category rather than in observations per category.

Table 10

Range of Mean Values for Teacher Observations
(Values in Number of Minutes)

Content Categories	Minimum	Maximum	Behavior Categories	Minimum	Maximum
0	1.92	5.42	O	15.56	22.28
1*	2.13	6.75	A*	9.75	17.63
2	1.67	3.06	C	4.84	8.11
3	9.81	15.81	E	.72	1.25
4	1.47	5.97	F	.03	1.42
5	1.75	4.50	G	2.86	12.64
6	2.14	5.64			
7	9.64	16.22			
8,9,10	3.28	5.39			

* given for sections rather than for teachers

Clearly there was a broad variation in teaching styles used in these recitations. The most striking example is category G which monitors how much teacher-student interaction occurred. Clearly teacher patterns varied from one in which very little teacher-student interaction occurred (a lecture style), to one in which about 25% of a typical class period was devoted to teacher-student interactions.

3. Relations of Recitation Observations to Student Achievement.

As has been emphasized by Rosenshine and Furst,[3] it is important to try to relate the results of classroom observations to student outcomes such as course achievement. This step is important so that the field of education can eventually accrue evidence as to the value of various teaching characteristics, both cognitive and affective. At present only a small number of affective characteristics and even fewer cognitive characteristics have consistently been found to be related to student achievement.

3. B. Rosenshine and N. Furst, "The Use of Direct Observation to Study Teaching", in Second Handbook of Research on Teaching, ed. by R.M.W. Travers, (Chicago: Rand McNally College Publishing Co., 1973), p. 122.

In considering the relationships between teacher behaviors and student achievement, one must keep in mind the limits of the present research. 1) First, a small group of teachers was involved in the study seven teachers in fourteen recitation sections. 2) Recitation was only one instructional component of the course. Students with unsatisfactory recitation instructors could attempt to learn the course material through lecture, the learning center, texts, and so forth. 3) All results are correlational rather than causal. That is, this study did not attempt to alter teacher behavior and then to determine differences in student achievement. 4) Our measure of student achievement was fairly narrow. Due to possible differences in grading practices for the recitation portion of the grade, and to the fact that lab work was fairly independent of recitation work, we used exam grades as the measure of achievement in this part of the analysis.

Within these narrow bounds we analyzed the correlation between teacher behaviors and student achievement.

As the measure of teacher behavior we used the total number of observations in each observation category for each recitation section. The totals are given in Table 11 for audio-tutorial sections and in Table 12 for standard sections. In Table 11 all totals have been multiplied by a factor of two since these sections were observed half as often as were the standard sections.

As the measure of student achievement, we used exam grades, adjusted for SAT math scores and math pretest scores. For each student we formed a total exam grade from the average of the two prelim grades plus the student's final exam grade. The average value of this total exam grade for each recitation section is given in the first column of Table 13. Using analysis of covariance, we obtained adjusted values of the total exam grade for each recitation section. The covariates were the SAT math scores and the math pretest scores. The adjusted total exam grade, the average SAT math score, and the average math pretest score are reported for each section in Table 13.

Table 11

Observation Category Totals

for Audio-Tutorial Recitation Sections*

Teacher: Category	I	II	III	IV	V	VI	VII
0	56	246	110	116	118	212	148
1	154	162	52	58	82	148	66
2	100	58	40	38	70	46	82
3	254	220	340	362	350	170	318
4	44	104	148	74	126	100	118
5	56	96	46	24	40	42	26
6	56	52	100	104	72	164	100
7	354	184	286	294	306	348	284
8,9,10	66	98	80	56	78	66	36
0	524	598	396	426	376	604	352
A	342	234	360	288	298	372	288
C	162	166	122	170	174	92	124
E	16	18	28	16	16	20	16
F	6	0	6	0	2	2	26
G	72	182	252	220	316	188	346

* All totals have been multiplied by two in order to make these totals comparable to those for standard sections which were observed twice as often as audio-tutorial sections.

Table 12

Observation Category Totals

for Standard Recitation Sections

Teacher Category	I	II	III	IV	V	VI	VII
0	41	49	67	83	136	86	85
1	51	74	98	57	57	89	75
2	33	48	40	59	68	87	40
3	346	357	363	313	298	268	346
4	31	121	112	51	129	104	156
5	48	84	57	77	50	42	53
6	49	81	64	119	83	121	85
7	407	255	221	360	255	311	282
8,9,10	143	145	95	90	94	92	110
0	443	463	365	475	394	500	384
A	423	315	296	421	267	352	344
C	189	209	143	145	182	128	142
E	21	18	31	23	20	29	18
F	3	14	1	1	3	4	38
G	67	170	244	134	268	171	282

Due to the fact that the same teachers taught one audio-tutorial recitation and one standard recitation section, we analyzed the data separately for the audio-tutorial sections and for the standard sections and then compared the results of these two analyses. Thus for each set of seven recitation sections and for each observation category we obtained the correlation coefficient assessing the degree of relationship between total numbers of observations and adjusted exam totals. These coefficients are listed in columns one and two of Table 14. Finally, we averaged the two correlation coefficients for each observation category using Fischer's r to z transformation. These average correlation coefficients are listed in the third column of Table 14.

The correlation coefficients listed in Table 14 indicate trends in the data. Of course due to the small number of teachers involved these trends are only suggestive. Six correlation coefficients had the same sign in both groups of recitation sections and were moderately large. Two categories, 1 and 0, were negatively correlated with adjusted exam scores. Four categories, 3, 4, C, and G, were positively correlated with exam scores. These data suggest a negative relationship between an emphasis on course and classroom administration (1) and student achievement on exams. Low usage of behaviors listed in the behavior categories was also negatively correlated with student achievement. This type of teaching would be categorized in behavior category 0 and would be characterized by a large amount of lecturing without props, writing on the blackboard, or interaction with students. On the other hand, describing a physical system or process or a demonstration (3), stating a problem or rephrasing a problem (4), drawing a diagram on the blackboard (C), and questioning or discussion with one student (G) were each positively correlated with student achievement. Note that categories 3 and C often described the cognitive and behavior characteristics of a given teacher behavior, as were the categories 4 and G. Thus the four positive correlations found here are probably reducible to two -- between the use of behavior which would be categorized as 3C and the use of behavior that would be categorized as 4G. Similar remarks apply to categories 1 and 0, since these often were used together to categorize a given teacher behavior.

BEST COPY AVAILABLE

Table 13

Mean Achievement and Trait Data
for each Recitation Section

	Teacher	Number of Students*	Total Exam Score**	Adjusted Total Exam Score**	Math SAT	Math Pretest Errors
Audio-tutorial sections	I	26	120.92	119.48	690	2.42
	II	14	96.32	97.91	687	3.07
	III	21	131.95	126.42	694	1.81
	IV	18	107.39	114.83	661	3.67
	V	19	125.08	127.74	679	3.11
	VI	22	112.43	110.23	691	2.41
	VII	18	111.28	110.52	690	2.67
	Teacher					
Standard sections	I	17	108.06	112.59	689	3.71
	II	25	129.04	129.37	672	2.48
	III	27	116.22	118.31	669	2.78
	IV	23	119.65	118.18	698	2.70
	V	19	127.05	126.67	683	2.58
	VI	18	113.53	111.58	698	2.61
	VII	19	117.61	115.30	695	2.47

* The same criteria were used for including students' data as were stated on page 20 of this report.

** Scores differed significantly across the fourteen sections (level of significance = .05.)

Table 14

Correlation Coefficients for
Adjusted Total Exam Scores vs Observation Categories

Category	Audio-tutorial (n=7)	Standard (n=7)	Average*
0	-.788	.236	-.397
1	-.545	-.143	-.354
2	.055	-.011	.025
3	.612	.224	.438
4	.224	.376	.300
5	-.514	.617	.081
6	-.010	-.098	-.055
7	.571	-.564	-.064
8,9,10	-.085	.191	.000
0	-.619	-.170	-.422
A	.548	-.611	-.046
C	.086	.663	.414
E	.284	-.441	-.092
F	-.031	-.025	-.025
G	.202	.311	.254

* Averages were formed using Fischer's r to z transformation

As discussed earlier, this analysis has used a fairly narrow definition of student achievement, i.e., exam grades, and has ignored other possible measures of physics achievement. In particular it has ignored the development of higher order cognitive abilities such as the ability to generalize or to specialize. Development of these higher order cognitive abilities might be associated with emphasis on concepts or theories (category 5) or with generalizing or specializing (category 8, 9, 10).

4. Relation of Recitation Observations to Teacher Ratings

Finally, we attempted to relate student ratings of instructors to the instructors' use of the various content and behavior categories. We sought to determine whether any general pattern would emerge relating teaching behaviors or teaching styles to student ratings of teachers.* The ratings used were based on those developed and reported by Hildebrand, Wilson, and Dienst.[4] Students were asked to (anonymously) rate their recitation instructors on each of the following rating scales.

The teacher:

R1. Has command of the subject, contrasts various points of view, discusses current developments, and relates topics to other areas of knowledge.

R2. Makes himself clear, states objectives, summarizes major points, presents material in an organized manner, and provides emphasis.

R3. Is sensitive to the response of the class, encourages student participation, and welcomes questions and discussion.

R4. Is available to and friendly toward students, is interested in students as individuals, and is valued for advice not directly related to the course.

R5. Enjoys teaching, is enthusiastic about his subject, and makes the course exciting.

Ratings could run from 1 to 7, with 1 indicating that the items was "not at all descriptive of the teacher," and 7 indicating that the item was "very descriptive."

* Please note that our observation program focused on certain characteristics of classroom teaching and necessarily ignored other relevant data which may affect student ratings of teachers, such as the number of times the teacher was absent from class.

4. M. Hildebrand, R.C. Wilson and E.R. Dienst, "Evaluating University Teaching" (Berkeley Center for Research and Development in Higher Education, University of California, 1971), p. 23.

Results were analyzed by section rather than by teacher. Significant differences (at .01) were found for each rating scale. Mean values of sections' ratings are given in Table 15.

Sections were usually rated in about the same order on the five scales. The pattern was as follows:

From his a.t. section, teacher VI received the lowest or second lowest ratings of all teachers on all scales.

The a.t. section of teacher I and the standard section of teacher VII rated their teachers very highly. These two sections generally gave very high ratings on each scale.

As was the case with the analysis of observation-achievement relationships, our analysis of observation-ratings relationships was limited by the small number of sections studied. Nonetheless, a type of pattern did emerge in relation to teacher ratings.

Table 15

Mean Ratings of Recitation Instructors

	Teacher	Maximum Number of ratings received*	R1	R2	R3	R4	R5
Audio-tutorial sections	I	22	6.27	6.73	6.36	6.18	5.96
	II	13	6.08	6.08	6.00	5.50	5.69
	III	19	4.50	5.47	5.79	5.07	5.00
	IV	13	5.85	6.23	6.46	5.67	5.58
	V	10	5.50	5.20	6.30	5.70	4.80
	VI	19	4.63	4.95	4.58	4.12	4.17
	VII	20	5.65	5.45	6.15	5.47	5.20
Standard sections	I	12	5.75	6.25	6.25	5.92	5.83
	II	28	5.61	5.22	5.07	3.92	4.52
	III	24	5.09	5.54	5.75	5.09	5.13
	IV	19	5.00	5.95	5.47	5.00	4.74
	V	12	5.08	5.75	6.33	5.17	5.42
	VI	14	4.93	5.50	6.14	4.85	4.93
	VII	16	6.44	6.19	6.31	6.00	5.81

* Total numbers of responses differed for each rating. For R1, N=238; R2, N=239; R3, N=241; R4, N=223; R5, N=237.

The two teachers receiving the highest ratings, teacher I and teacher VII, were quite different from each other in style. Teacher I featured a lecture style emphasizing problem solutions (7) and teacher VII featured a questioning style (G) emphasizing problem statement (4).

On the other hand, the teacher who received the lowest ratings, teacher VI, was in the middle on category 4, category 7, and category C. He was high on category 6 which indicates an emphasis on mathematical explanations of problems and 0 which indicates that he simply talked without using any written support. Finally, he was lowest among teachers on categories 3 and C which refer to the description of physical phenomena and the use of drawings on the blackboard.

These results, although quite sketchy, appear to indicate that the teachers who used highly developed teaching styles were rated significantly higher than was the teacher who used a fairly undeveloped style of teaching.

V. Conclusion: Guidelines for Course Decisions

This research has been conducted in part to assist in the development of guidelines for course decisions. These decisions are in the following areas:

1. methods of instruction;
2. procedures for assigning students to methods;
3. coordination of methods of instruction;
4. the operation of the learning center;
5. selection of teaching styles for recitation.

As far as selection of a method of instruction is concerned, both methods lead to similar levels of achievement for the majority of randomly assigned students. However, students with relatively low math achievement and math aptitude may profit more from the audio-tutorial method than the stand method. Conversely, students with very high math achievement and mat tude may profit more from the standard method than from the audio-t orial method.

If both methods of instruction are to be used in a course such as P112, one should keep these factors in mind in the selection of students. If only one method is to be selected for all the students in a course, the students' math achievement and math ability characteristics ought to be major criteria in the selection of a method of instruction.

In regard to selection of a procedure for assignment of students to methods, assignment by preference has many drawbacks in relation to random assignment. It is more difficult to assign by preference. When preference is used as an assignment procedure, some students are assigned contrary to their preference so as to balance class sizes. This leads to confusion and dissatisfaction. When preference is the assignment procedure, students who select the audio-tutorial method appear to be overconfident in regard to their success in the course. They under-utilize the available materials and achieve significantly less than do students who choose the standard method.

However, students assigned by preference tend to have more favorable attitudes to the method of instruction they receive than students who are randomly assigned. Nevertheless, students who are randomly assigned generally remain in the method to which they are assigned.

On balance, random assignment appears to be preferable to assignment by preference, particularly in the light of the consistently lower levels of achievement obtained by the a.t. preference group, the failure of a.t. preference students to utilize the method they rate highly, and their failure to compensate for lower performance in P112 by performing better in their other courses.

If both methods of instruction will be used in a course, the best assignment procedure would seem to be as follows:

1) assign students at the upper end of the math ability and math achievement range to standard instruction;

2) assign students at the lower end of the math ability and math achievement range to audio-tutorial instruction;

3) randomly assign students in the middle range.

In regard to the coordination between the two methods of instruction, attention is needed in order to avoid differing emphases and coverage in the two methods. Such differences result in student and teacher dissatisfaction.

In regard to the learning center operations in P112, some difficulties are apparent. Observations indicate that periodically there were very heavy loads on the lab facilities of the learning center, causing breakdown of apparatus and overflow into other facilities. These problems should be alleviated by increased technical assistance in the lab. Furthermore, certain instructors working as tutors interact little if at all with students in the learning center. Due to the fact that there is very little student-to-student interaction in the learning center, some effort is needed on the part of tutors to interact with students and thereby to personalize teaching in the learning center.

In regard to recitation instruction, observations support the following:

Less recitation time should be devoted to quizzes in the audio-tutorial sections than in the standard sections. This would remove some of the time pressure felt by the audio-tutorial teachers and might result in increased discussion at higher order cognitive levels.

If recitation instructors are interested in improving student attitudes toward their teaching, they might be advised to develop a definite teaching style. They might, for instance, emphasize either exposition of material or questioning and discussion. The teaching style which is developed should include adequate use of written supports such as equations and diagrams.

Appendix A

A TRAIT-TREATMENT INTERACTION IN A COLLEGE PHYSICS COURSE

[To be published in the

Journal of Research

in Science Teaching.]

Mary Diederich Ott

School of Applied and Engineering Physics
Department of Physics

and

Center for Improvement of Undergraduate Education

Cornell University
Ithaca, New York

and

David B. Macklin

Center for Improvement of Undergraduate Education

Cornell University
Ithaca, New York

Introduction

Frequently, comparisons of student achievement in two or more instructional treatments are reported to result in no significant differences between treatments[1]. However, as has been emphasized recently[2], in such studies one also ought to look for interactions between student traits (including aptitudes) and treatments (e.g., instructional methods). If such interactions are found, one treatment may be better for some students having certain levels on the trait in question, and the other treatment may be better for students having other levels on that trait. It is clearly important to identify such interactions so as to guide faculty and students in selecting instructional methods.

We have used a trait-treatment interaction approach to analyze the achievement of first year college physics students in two instructional treatments -- an audio-tutorial treatment, and a lecture-recitation-laboratory treatment. We have found indications of a disordinal* interaction between the treatments and mathematical aptitude (as measured by the mathematics portion of the Scholastic Aptitude Test) and between the treatments and mathematical achievement (as measured by a pretest of our design). The dependent variable was the students' achievement in the course as measured by course final grade.

* A disordinal interaction is one in which regression lines for the two treatments cross within the observed range of the measured trait.

The Course and the Treatments

The study was conducted in reference to a one-semester freshman level physics course offered to approximately 575 engineering and physics majors at Cornell University. About 98% of the students were freshmen, 91% were enrolled in engineering, and 96% were males. Most students were taking the course because it was required. The subject matter of the course included classical mechanics, special relativity, and heat. One semester of calculus was a prerequisite for registration in the course. The course textbooks were Fundamentals of Physics by David Halliday and Robert Resnick and Space and Time in Special Relativity by N. David Mermin.

The course was offered to students via one of two treatments -- a lecture-recitation-laboratory (standard) method of instruction and an audio-tutorial method.[3] The content of the instruction was very similar in the two treatments. In addition students in both treatments had the same homework assignments, closely similar laboratory assignments, and identical examinations. The difference in treatments, therefore, was primarily one of instructional method, rather than of content, assignments, or examinations.

The standard method of instruction included two hours of lecture and two hours of recitation per week, and a two hour lab every other week. (Recitation refers to a small group session used for problem solving, discussion, and quizzes.) There were lecture sessions at two scheduled hours twice a week, with about 125 students assigned to one of these sessions, and about 300 assigned to the other session. On the average, 26 persons were assigned to each recitation section, and about 12 to each lab section.

The lectures were taught by Prof. John Silcox, who also was the co-author of materials used in the audio-tutorial treatment. Recitations were taught by faculty members and graduate teaching assistants. Laboratory sections were taught by graduate teaching assistants.

The audio-tutorial method included one hour of recitation per week. This hour of group instruction was included primarily to provide for group interaction, as well as student contact with one particular instructor and an opportunity for testing. In the audio-tutorial method, all other instruction took place at the student's convenience in a learning center which was staffed by tutors (i.e., physics graduate students and faculty members) fifty-two hours per week. Materials available in the learning center included demonstration apparatus for self-demonstrations, laboratory equipment, audio-tape commentaries, and slides. The taped commentaries and slides were coordinated with a study guide developed by Dr. OTT and Professor Silcox. Tape-slide-study guide sequences were used in particular to assist students in learning to solve physics problems.

Research Design

Each enrolled student was assigned to one of the twenty-two recitation sections, scheduled at eight different class hours for meetings twice a week. Students were assigned to a section at a given class hour on the basis of their other scheduled classes and personal preference. Fifteen of the twenty-two recitation sections scheduled at four of the eight class hours were selected for inclusion in the study. These four class hours were selected because three or more sections were scheduled at each of these hours.

In order to study the effects of student preference for each of the methods of instruction, we used two procedures of assignment to treatments. At two of the four class hours included in the study, students were randomly assigned to treatments and to recitation sections within the treatments. At the other two hours, students were assigned to treatments on the basis of their preferences as expressed on a questionnaire administered at class meetings during the first week of the semester. Within each hour, students were then randomly assigned to specific sections within the preferred treatment.* Table 1 illustrates the distribution of treatments, procedures of assignment, and instructors among the fifteen recitation sections.

---

* At each of these hours, more students chose the standard method of instruction than could be accommodated if section sizes were to be balanced. Therefore some students who indicated that their preference for the standard method was not strong were assigned to the audio-tutorial sections. In all there were 24 students reassigned in this manner, an average of 8 to each audio-tutorial preference section. Subsequently six of these students joined standard sections. These six students were not included in the data analysis.

Enrollment within the fifteen sections varied from 18 to 28. 150 students in six sections were taught by the audio-tutorial method; 242 students in nine sections were taught by the standard method. Data on only 303 students (115 in the audio-tutorial treatment and 188 in the standard treatment) were analyzed. Of the other 89 students, 78 either dropped the course or transferred to other sections of the course; 11 had incomplete data.

In order to study trait-treatment interactions without confounding treatment effects and traits, one needs to measure students' aptitudes or traits prior to or at the very beginning of the treatments. In this study, traits were (1) reported by students on a questionnaire, or (2) measured by means of a math pretest of our design*, all completed at the first meeting of the recitation sections. From the available information, the choice of traits to be tested for interactions with treatments was based on the following criteria:

(1) importance of the trait with respect to the course and to the differences in the treatments;

(2) reliability and validity of the measurement of the trait;

(3) range of responses to the trait measure.

Of the possible traits determined by the questionnaire and the pretest, the following three met the criteria:

(a) mathematical aptitude, as measured by the Mathematics portion of the College Entrance Examination Board Scholastic Aptitude Test (SATM);

(b) verbal aptitude, as measured by the Verbal portion of the Scholastic Aptitude Test (SATV)**,

(c) mathematics achievement, as measured by the math pretest.

---

* The nine-item multiple choice math pretest is a shortened version of a twenty-item pretest used in the same course during the Spring semester, 1972. The latter was shortened by eliminating items that did not discriminate among different levels of mathematical achievement, i.e, those items which were answered by almost all or almost none of the students. In contrast to SATM items, pretest items were selected because they were representative of the mathematical skills required of the students in this specific course. Within each treatment, the correlation between math pretest scores and SATM was equal to .33. (Please see Table 4.) Thus the portion of the variance in math pretest scores accounted for by SATM levels is only 11%, and the SATM and Math pretest appear to have measured different student traits.

** Both SATM and SATV were coded in 13 levels.

The correlation coefficients of each of these traits with final grade in each treatment are listed in Table 2. A multiple regression analysis of the relation of the three traits to the course final grade indicated that SATV was not significantly related to final grade in either treatment. Thus SATM and math pretest scores were found to be the best predictors of final grade in each treatment. The subsequent analysis considers only these two traits.

All students were required to take six scheduled quizzes, generally one quiz every other week, as well as two interim examinations and a final exam. Testing thus was not on a self-paced or a mastery basis in either treatment. Quizzes were constructed by the senior author (Dr. Ott), and exams were constructed by committees of recitation, laboratory, and lecture instructors.

At the end of the semester a course grade was calculated for each student. The grade was determined as follows:

Each student received

(a) 0-100 points for lab (based mainly on lab reports);

(b) 0-100 points for recitation (based mainly on quizzes);

(c) 0-100 points for the average of the two interim examinations;

(d) 0-100 points for the final examination. *

The total scores thus could range from 0 to 400. Grades were allocated in such a way that the median course grade was C+, equivalent to 235 to 249 total points.

The course final grade therefore reflects achievement on lab reports, quizzes and exams, and performance in lab and recitation sections. As such, it is a broad-gauge measure of achievement in the course. For this reason the course grade was used as the measure of achievement (dependent variable) in this study.

---

* The variances of these four subscores were as follows: (a) 52; (b) 315; (c) 339; (d) 311. Thus lab subscores had less effect than the other subscores on student rankings and final course grades.

Comparisons of initial trait measures and of final grades yielded the results given in Table 3. The differences in traits which are significant at .05 are the following:

(1) Within the random assignment sections, the variance of SATM levels is significantly larger in the audio-tutorial treatment than in the standard treatment.

(2) For all sections, the variance of SATM levels is significantly larger in the audio-tutorial treatment than in the standard treatment. The mean value of math pretest scores is significantly larger in the audio-tutorial treatment than in the standard treatment.

There are no significant differences in the means or variances of final grades.

Evidence of an interaction between the traits (math pretest score and SATM level) and the treatments is shown by the differences between the regression coefficients in the two treatments. The regression equations for the two treatments are as follows (X is the score on the math pretest, Y is the level on SATM, Z and U are estimated grades in the audio-tutorial and standard treatments, respectively):

Audio-tutorial: $Z = -.5175\ X_1 + .2533\ Y_1 + 9.1667$

Standard: $U = -.8495\ X_2 + .4081\ Y_2 + 10.6352$

Note that the coefficients of X_1 and of X_2 are negative because grades are coded 1 to 13 from high to low and pretests are coded 1 to 9 from low to high.

In order to determine whether there are levels on either or both of the traits which predict greater achievement in either treatment (and thus to determine whether the interaction is significant), we analyzed the interaction using the Johnson and Neyman technique*. We used the notation given by Johnson and Neyman[4], and in an instructive article by Koenker and Hansen[5]. The Johnson-Neyman technique allowed us to determine the conjoint math pretest scores and SATM levels for which it was reasonable to assume that the difference in grades between treatments was real. The basic statistics are listed in Table 4.

---

* Note that if one were to obtain significantly different coefficients for a single trait, one would have evidence of an interaction between that trait and the treatments. In that case it would not be necessary to use the Johnson-Neyman technique to determine regions of significant differences.

We set the level of significance at .10, and defined

$$w_{.10} = \frac{N_1 + N_2 - s}{F_{.10}}$$

, where s is the number of parameters.

In this case $w_{.10} = (303 - 6)/\ 2.74 = 108.39$. We obtained an observed value of the test statistic w_{obs} equal to 74.86. Thus since $w_{obs} < w_{.10}$, a region of significance exists at the ten percent level. The region in the x', y' plane is hyperbolic and consists of two parts, as indicated in Figure 1. At the .10 level of significance, the standard treatment is preferable to the audio-tutorial treatment in terms of expected achievement for students with math pretest scores equal to 9 and SATM greater than or equal to 725. The audio-tutorial treatment is preferable to the standard treatment for students with math pretest scores less than or equal to 4 and SATM less than 625. Both methods of instruction are equivalent in terms of final grades for students whose scores do not fall within these limits.

Discussion

The foregoing analysis has indicated an interaction between two traits (scores on a math pretest and levels on SATM) and the method of instruction, when the dependent variable is course grade. A significant question remains, however. What aspects of the differences between treatments account for the interaction?

As we have indicated, the two treatments used different instructional components. The standard method included two hours of lecture and two hours of recitation per week and a two hour lab every second week. The audio-tutorial method included one hour of recitation per week in addition to learning center activities.

Thus, there were major differences between the treatments in instructional components and in the amount of time allocated to each instructional component. These differences include:

(1) An emphasis on lecture (large group instruction) in the standard method of instruction; an emphasis on learning center activities (one-to-one instruction) in the audio-tutorial method.

(2) A greater amount of recitation time (small group instruction) in the standard method.

(3) Group laboratory activities in the standard method; individual laboratory work in the audio-tutorial method.

There are other differences between the treatments in addition to those related to major instructional components. These differences include:

(4) A greater degree of individualization in the audio-tutorial treatment, i.e.,

(a) students could obtain instruction when they wished for as long as they wished;

(b) students could stop and repeat instruction;

(c) students could obtain immediate tutorial assistance.

(5) A greater amount of active rather than passive learning in the audio-tutorial treatment.

(6) The availability of a wider variety of instructional media to students in the audio-tutorial treatment.

(7) The coordination of laboratory activities with other activities in the audio-tutorial treatment.

(8) An emphasis on step-wise solution of problems in the audio-tutorial treatment.

(9) Development of notes in the audio-tutorial study guide suitable for review.

Additionally, there may have been differences which are not intrinsic to the two methods of instruction. For instance, it is possible that in these particular realizations of the treatments, the treatments were interesting to or suitable for students of different abilities or levels of achievement. That is, one treatment may have been better for higher ability or achievement students and the other for lower ability or achievement students. In that case, the interaction might be due to characteristics of the particular realizations of the treatments rather than of the treatments themselves. If this were the case, one would expect students who were not well-matched in their instructional method to stop attending the lecture or learning center while students who were well matched to continue in attendance. Thus, for instance, if a greater percentage of students from the lower ability (SATM) levels stopped attending a treatment than students at the higher ability (SATM) levels, one might suspect that this treatment was less suited to the lower ability students.

In this study, we found that there were no major differences in continued attendance in lecture among students having different scores on the SATM. However, there was some tendency, though not strong, for students with higher scores on the math pretest to attend a greater percentage of lectures than those with lower scores on the pretest. There was no relation between continued learning center attendance and SATM or math pretest scores. The observed interaction is, therefore, probably due to differences inherent in the treatments, e.g., those labeled (1) to (9) above.

Of these nine items, those are of interest which tend to account for the fact that correlations between the math pretest and achievement (and between SATM and achievement) are lower in the audio-tutorial section than in the standard section. Thus we are interested in those items which tend to obviate differences in mathematical achievement and ability among the audio-tutorial students. In particular, we are interested in those differences between treatments which (1) give an advantage to high ability students in the standard treatment relative to those in the audio-tutorial treatment; or (b) give an advantage to low ability students in the audio-tutorial treatment relative to those in the standard treatment.

Of the first set of three differences discussed above, (1) and (3) might tend to benefit lower ability and achievement students in the audio-tutorial treatment. This would result due to the greater individualization of the learning center and audio-tutorial lab in comparison with lecture and the standard lab. On the other hand, point (2) would seem to favor students in the standard treatment because these students had twice as much time devoted to recitation. Because recitations generally emphasize problem solving, it is reasonable to assume that they might have been most beneficial for the high ability or achievement students. If this were the case, point (2) would help to account for the superiority of these students in the standard treatment relative to those in the other treatment. A measure of support for this assumption comes from the fact that in the standard treatment the students having higher math pretest scores attended a higher percentage of recitations than those with lower math pretest scores. Thus, recitation may have been particularly useful for high achievement students.

Of items (4) through (9), all might be particularly helpful for those students in the audio-tutorial treatment with lower aptitude and achievement levels for the following reasons. Items (4) through (9) indicate that students in the audio-tutorial treatment could spend the time required on portions of the subject matter which caused them particular difficulty. Moreover, students could work in a step-by-step fashion, and could engage in as much review as they found necessary. On the other hand, the kind of instructional assistance described in these items would be of less importance to the highest ability students in the audio-tutorial treatment. Therefore, these students would not have an advantage relative to similar students in the standard treatment on the basis of items (4) through (9).

We conclude, therefore, that the observed interaction is probably due to inherent characteristics of the two treatments, such that one was more suitable for students in our population with the highest ability and achievement, and the other more suitable for the students in our population with the lowest ability and achievement. Recall that the range of SATM scores was from about 500 to 800. The greater degree of individualization in the audio-tutorial treatment seems to benefit lower ability and achievement students relative to similar students in the standard treatment. The greater amount of time allocated to recitation in the standard treatment, and the emphasis in recitation on problem solving, may account for the greater achievement of higher ability and achievement students in the standard treatment relative to similar students in the audio-tutorial treatment.

These results are in general congruent with Cronbach and Snow's interpretation of earlier trait-treatment interaction studies. As summarized by Salomon[6], treatments which force students to pay attention to detail appear to benefit low general-ability students, in relation to higher ability students. On the other hand, treatments requiring rapid manipulation of symbolic meaning appear to benefit the higher ability students rather than low ability students.

At present, it is impossible for us to generalize from the interaction which we have observed in a study of two treatments in one physics course to other treatments and other courses. * It would be most useful in

---

* This research is continuing under a grant from the Sloan Foundation.

this regard if other investigators would review their research comparing instructional treatments to determine whether they have evidence of possible trait-treatment interactions.*

Synopsis

We have used a trait-treatment interaction approach to analyze the achievement of college physics students in two instructional methods--an audio-tutorial method of instruction and a lecture-recitation-laboratory method. We have found indications of an interaction between mathematical aptitude, achievement in mathematics, and the instructional methods. The dependent variable was the students' achievement in the course as measured by course final grade.

The lecture-recitation-laboratory method appears to be more suitable for students with the highest mathematical ability and achievement, the audio-tutorial method more suitable for students with lower mathematical ability and achievement. Both methods of instruction are equivalent in terms of final grades for students between these two extremes.

---

* We thank Walter Federer and Jason Millman for helpful suggestions.

REFERENCES

1. R. Dubin and T. C. Taveggia, The Teaching-Learning Paradox: A Comparative Analysis of College Teaching Methods (University of Oregon Press, Eugene, 1968), pp. 24-35.

2. D. C. Berliner and L. S. Cahen, Review of Research in Education 1, 58 (1973); L. J. Cronbach and R. E. Snow, "Individual Differences in Learning Ability as a Function of Instructional Variables" (Stanford University), pp. 1-25.

3. For a general description of the original form of audio-tutorial instruction see S. N. Postlethwait, J. D. Novak, and A. T. Murray, Jr., The Audio-Tutorial Approach to Learning (3rd ed. Burgess, Minneapolis, 1972).

4. P. O. Johnson and J. Neyman, Statistical Research Memoirs 1, 57 (1936).

5. R. H Koenker and C. W. Hansen, Journal of Experimental Education 10, 164 (1942).

6. G. Salomon, Review of Educational Research 42, 327 (1972).

Figure Legend

The joint frequencies of SATM and pretest scores for each treatment are indicated at each point (x', y'). The number of data points (i.e., the joint frequencies) for the audio-tutorial treatment are listed at the upper right at each point (x', y'), and those for the standard treatment are listed at the lower left at each point (x', y'). The SATM levels are indicated on the y axis, with their respective scores indicated in parentheses. The math pretest scores are indicated on the x axis. The scores of students highest in mathematical aptitude and achievement as measured by the two traits are located in the upper right hand corner. The scores of students lowest on both traits are located in the lower left hand corner. The line $O\xi$ running diagonally from lower left to upper right is the axis of the hyperbolic region of significance. The $O\eta$ line intersecting the $O\xi$ line is the line of non-significance. On the upper right side of the $O\eta$ line the value of the difference in predicted grades (i.e., Z-U) is at least marginally in favor of students in the standard treatment. On the lower left side of the $O\eta$ line the difference in predicted grades is in favor of students in the audio-tutorial treatment. The portion of the hyperbola in the upper right corner is the region in which students in the standard treatment can be expected to achieve higher grades than students in the audio-tutorial treatment, at the .10 level of significance. This region includes students having math pretest scores of 9 and SATM scores of 725 and above. The portion of the hyperbola in the lower left corner is the region in which students in the audio-tutorial treatment can be expected to achieve higher grades than students in the standard treatment, at the .10 level of significance. This region includes students having both pretest scores less than or equal to about 4, and SATM scores lower than 625.

BEST COPY AVAILABLE

TABLE 1

Distribution of Treatments, Assignment Procedures, and Instructors Among Fifteen Recitation Sections.

Assignment Procedure of Student to Section	Class Meeting Time	Recitation Instructors* Audio-Tutorial	Standard
Random	9:05	A, B	C, D, E
Random	12:20	F	G, C
Student Preference	10:10	H, G	I, F
Student Preference	11:15	I	H, A

*Instructors A, C, F, G, H, and I taught two sections each. B, H, and I are faculty members; the remainder are graduate students.

TABLE 2

Correlation Coefficients of Traits with Final Grade, Within Treatments.

Trait	Audio-Tutorial Treatment (n = 110)	Standard Treatment (n = 172)
SATM	.28	.37
SATV	.18	.09
Math Pretest*	-.33	-.47

*The negative value of the correlation coefficient is due to the fact that math pretest scores were coded from low to high, whereas the grades were coded from high to low.

TABLE 3

Means and Standard Deviations of Initial Traits and Final Grades

BEST COPY AVAILABLE

Measure	Statistic	Within Assignment Procedure: Random		Within Assignment Procedure: Preference		All Sections	
		Audio-tutorial	Standard	Audio-tutorial	Standard	Audio-tutorial	Standard
		(n=57)	(n=101)	(n=58)	(n=87)	(n=115)	(n=188)
SATN	mean	5.21	4.71	4.55	4.78	4.88	4.74
	s.d.	2.82*	2.16	2.36	2.23	2.57*	2.19
Math Pretest	mean	6.93	6.43	6.91	6.56	6.92*	6.49
	s.d.	1.58	1.62	1.58	1.94	1.57	1.77
Final Grade	mean	6.44	7.25	7.29	6.84	6.87	7.05
	s.d.	3.39	3.58	3.32	3.82	3.37	3.69

*Differences significant at the 5% level, two-tailed.

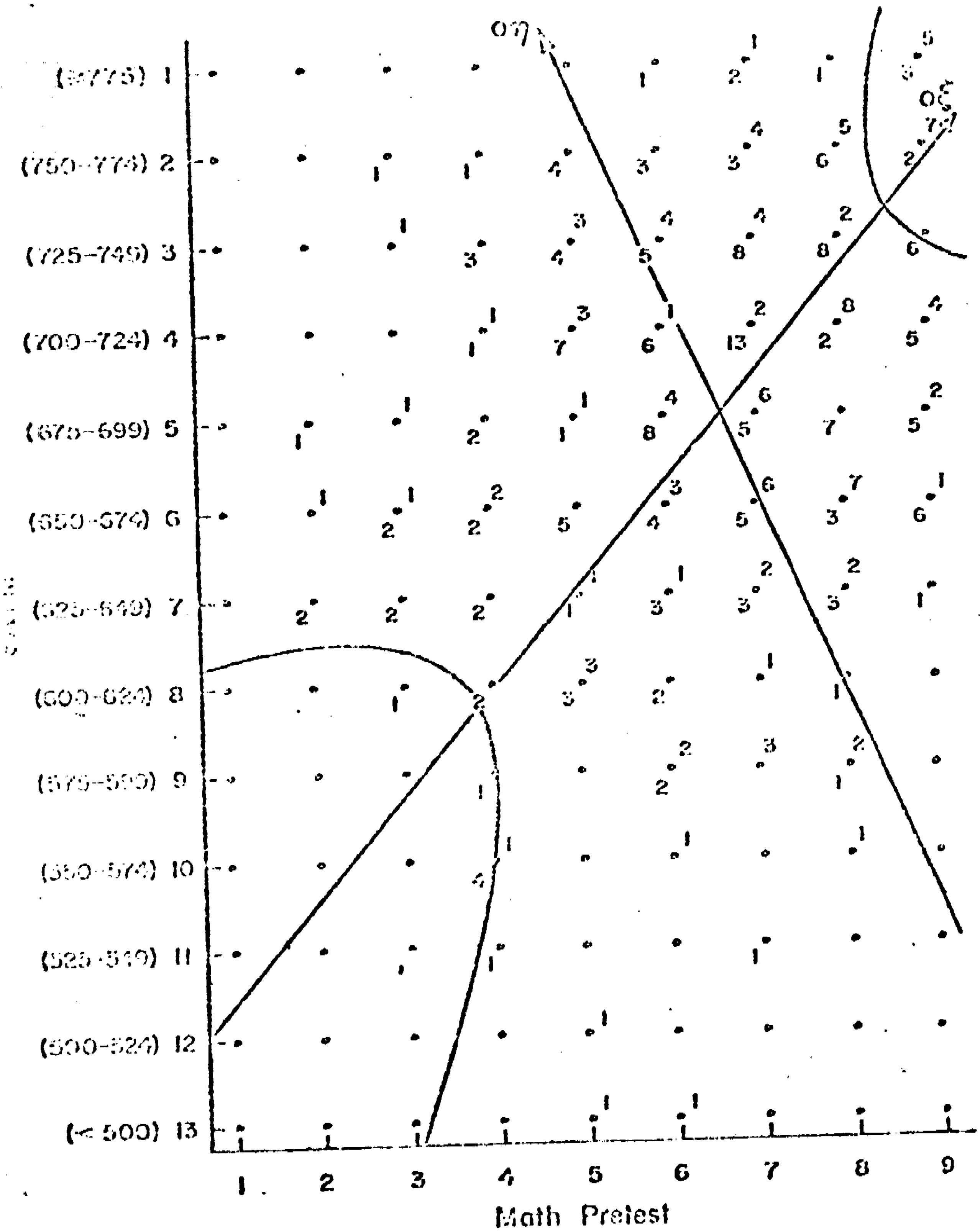
BEST COPY AVAILABLE

TABLE 4

Basic Statistics for Johnson-Neyman Analysis*

	Treatments: Audio-tutorial	Treatments: Standard
	$N_1 = 115$	$N_2 = 188$
Means	$\bar{x}_1 = 6.9217$	$\bar{x}_2 = 6.4894$
	$\bar{y}_1 = 4.8783$	$\bar{y}_2 = 4.7447$
	$\bar{z} = 6.8696$	$\bar{u} = 7.0585$
Standard Deviations	$\sigma_{x'} = 1.5736$	$\sigma_{x''} = 1.7717$
	$\sigma_{y'} = 2.6095$	$\sigma_{y''} = 2.1911$
	$\sigma_z = 3.3679$	$\sigma_u = 3.6916$
Correlation Coefficients	$r_{xy}' = -.3335$	$r_{xy}'' = -.3299$
	$r_{xz} = -.3098$	$r_{xu} = -.4876$
	$r_{yz} = +.2846$	$r_{yu} = +.3767$

* X = Math Pretest Score; Y = SATM level; Z = Final Grade in the audio-tutorial treatment; U = Final Grade in the Standard treatment.


(>775) 1
(750–774) 2
(725–749) 3
(700–724) 4
(675–699) 5
(650–674) 6
(625–649) 7
(600–624) 8
(575–599) 9
(550–574) 10
(525–549) 11
(500–524) 12
(< 500) 13
1
2
3
4
5
6
7
8
9
Math Pretest

BEST COPY AVAILABLE

Appendix B: Content Categories

0 = Silence, including silence during quizzes and films

1 = Course and classroom administration
(actually giving assignments, collecting homework, calling class to order, handing out papers, asking for quiet, discussing grading of exams, etc.)

2 = References to content of assignments, exams, other parts of course such as lecture or lab
(to content of a prior or upcoming exam or homework assignment; to material covered in lecture [recitation]; to lab work; to a.t. materials; brief reference to topics of upcoming lectures or recitations [naming the topics]; giving answers to problems without explaining them.)

3 = Describing a physical system or process or a demonstration; using a diagram or a graph to illustrate a process or kind of motion

4 = Stating a problem or rephrasing a problem

5 = Defining a concept, stating a theory, presenting a model, pointing out an analogy
(definitions may be presented in terms of words or equations)

6 = Applying mathematical techniques or operations without reference to physical implications or meaning
(performing a mathematical operation; deriving a mathematical concept or formula; correcting a mathematical error)

7 = Deriving or explaining an equation or formula with explicit reference to physical meaning or implications; discussing units
(discussion of meaning of a term in an equation or meaning of an equation; deriving a formula for a general case or a specific example, with explicit reference to the physics involved)

8 = Organizing or structuring the material
(transitional statements, e.g., saying what will be discussed and why; reviews or previews of topics)

9 = Beginning to specialize to a particular case from a more general treatment

10 = Beginning to generalize from a particular case to a more inclusive treatment

11 = Conversation only tangentially related to subject matter

Appendix C: Behavior Categories

0 = none of the others

A = writing an equation or words on blackboard

B = writing an equation or words on view-graph (or showing such a transparency on view-graph)

C = drawing a diagram on blackboard

D = drawing a diagram on view-graph (or showing such a transparency on view-graph)

E = erasing board, moving view-graph, or removing transparency from view-graph

F = performing a demonstration or showing a movie

G = questioning or discussion with one student

H = questioning or discussion with more than one student

I = presenting a question for use with Instant Response System

J = discussing student responses or correcting answer for Instant Response System question

Instructions for Use:

Notations are made every 30 seconds of the number corresponding to the two categories. During the first 15 seconds of each 30 second interval, observe the instruction and decide which category is represented. During the next 15 seconds write down the two category numbers.

Major topics and topic shifts are to be noted as they occur. That is, the name of a new topic is to be written down in the line of the topics column corresponding to the time at which the topic was introduced. Unusual events may also be listed here. Also, names of films, or descriptions of demonstrations should be listed here.

The category labelled "Content" refers to the function of the content being presented by the teacher (or by the student, during a time interval in which student talk is dominant.)

The category labelled "Behavior" primarily refers to non-verbal behavior but is also used to indicate discussion or questioning behaviors.

As ar as content categories are concerned:

-- write down 5, 7, 8, 9, or 10 any time they occur, even if others occur at the same time

-- otherwise, write down dominant content category.

As far as behavior categories are concerned:

-- write down I or J if they occur, even if something else also occurs

-- otherwise, write down the dominant behavior.

Appendix E: Observation Record

P.______

Date ______ Observer ______

Hour ______ Obs. started at ______

Teacher ______ Obs. ended at ______

Section # ______

BEST COPY AVAILABLE

C	B	TOPIC	C	B	TOPIC

Appendix F BEST COPY AVAILABLE

Studies of Observations of Teaching Related to Student Achievement

David B. Macklin

The purpose of this brief exploration is to select and recommend a strategy for the investigation of selections from tape recordings of recitation sections in a college physics course--the one described in the foregoing report. The tapes are from the same recitations rated by the observers (p.35).

The strategy has to meet two initial criteria: (1) It should be oriented to teaching behavior, because that is the predominant and important activity in the recitations; (2) The dimension or dimensions of teaching behavior attended to should be cognitive ones or related to such, because (a) cognitive learning is the primary aim of this course and (b) variations in teaching-learning among the thirteen sections may be expected.

A review of the major recent reviews of studies on observations systems related to students' learning was used as the primary method of search. The foremost review is Rosenshine and Furst's, in Travers (ed.), Second Handbook of Research on Teaching (1973). Rosenshine also did a review, which was somewhat different, that was published in Westbury and Bellack (eds.), Research into Classroom Processes (1971). Lastly, Nuthall has a major review in a 1970 Monograph of AERA. A partial check on coverage was made by using a computer search of the ERIC listings of 1968--mid 1974, for college level courses combined with observational studies of teaching. These reviews were scrutinized for studies which met the above criteria and were observations-based and related to student achievement. Within the search conducted, it might be noted, almost no observational studies of college teaching were found.

A few major studies will be summarized in order to discover the kinds of factors found relevant in the relationships of observational data and student learning. A study by Furst yielded a complex set of findings, from high school students in a 4-lesson economics "course". (See Nuthall, 1970, pp. 19-21.) The 15 classes involved yielded three achievement groups, relative to one another. The data analyzed were typescripts of the complete in-class talk. The findings, from Nuthall's summarization of Furst's thesis, were; (1) "...the three high-achieving classes differed from the others in having more 'extended' indirect teacher talk, more positive than negative immediate feedback to pupil responses, and more extended participation by the pupils." (2) A composite variable was constructed by Furst, incorporating (a) moderate vs. high or low "amount of verbal structuring moves", "a moderate rate of question-answer exchange", and "a high degree of variety in the kinds of logical processes exhibited." The high achieving group of classes differed significantly from the other two on this index. Another facet was brought out, by Rosenshine in the same AERA Monograph, to the effect that significantly more typescript lines were devoted to "defining" and "interpreting" in contrast to "fact-stating", by the teachers of the high achievement classes (Rosenshine, 1970, p. 116).

Two additional findings, based on the Flanders system (1965) of interaction analysis, were cited by Nuthall. They involve "lecturing" and what is labelled the I/D Ratio, standing for "Inditect" vs. "Direct" Teacher influence. The nature of the components of this ratio derive from summing the subclass entries in each of the two general categories:

Indirect: accepts feelings, praises/encourages, accepts or uses ideas of students, asks questions.

Direct: lecturing, giving directions, criticizing or justifying authority. (See Table 1 for the definitions.)

Furst (as reported in Nuthall, 1970, p. 20) found that the I/D ratio was related to class achievement, with the ratio being higher (= more Indirectness) in the high performance classes. These particular data point to a curvilinear relationship, because the low point is the "average: performance classes in association with the lowest I/D ratio. But replications are needed to test this, rather than higher powered statistical analysis of the one set of data. --Regarding lecturing, the lowest proportion occurred in the high performance classes (c. 30%), and the highest in the average group (c. 60%)

Wolfson (1973) reports a study of high school students in chemistry and in general science. Teachers in these classes who manifested greater indirectness (Higher I/D ratios) had taught the classes which did significantly better on major standardized tests. The same kind of finding was obtained by LaShier (reported by Nuthall, 1970, pp. 18-19) for 10 classes of 8th grade students in a 6-week lab unit from a biology (BSCS) curriculum; greater indirectness was significantly related to gains in students' learning and also to students' positive attitude toward the unit.

A study by Solomon, Bezdek, and Rosenshine, reported by Rosenshine in Westbury and Bellack (1971, pp. 65-67), also merits some attention. Solomon and his co-workers coded tape recordings, being the observations, to find differences which were related to two kinds of differences in student achievement: gain in factual knowledge and gain in comprehension of subject-matter. Twenty-four classes of a college level night school, one-semester course in American government were the study's population. Two sessions of each class were recorded. From these recordings, values for 61 measures or dimensions were taken. These were combined with 100 measures derived from ratings by observers and pupils. (It's reasonable to assume that rating is a summary depiction, in contrast to a specific time-and-place observation.) The 161 measures were factor analyzed into 6 factors, then related to the two types of gain: factual and comprehension. Four of the six factors yielded significant correlations with gain, but no factor was significantly related to both types of gain.

The results may be presented as follows, which shows both the factor label the authors used and the measure derived from the observations which was most closely related to that factor--in order to give a more concrete meaning to label.

Factor related to gain in factual knowledge:

"Clarity vs. Obscurity" (proportion of student requests for interpretation to total student speech [factor loading: -.56]) -- corelation (r) with factual gain = .58 ($p < .05$; linear relationship); r=.04 with comprehension gain (not significant).

Factors related to gain in comprehension:

"Permissiveness vs. Control" (proportion of teacher speech to total classroom speech [factor loading: -.92]) -- correlation was significant (though not reported numerically in Nuthall's review), but the relationship was curvilinear, i.e., middle levels of "Permissiveness

BEST COPY AVAILABLE

vs. Control" were associated with higher comprehension gains. The relationship to factual gain was insignificant.

"Energy vs. Lethargy" (ratio of teacher requests interpretation to student requests [factor loading: +.76]); r =.44 with comprehension gain ($p < .05$; linear); r =.23 with factual gain (not significant).

"Flamboyance vs. Dryness" (ratio of teacher personal references to student personal references [factor loading: =.51[); r =.42 with comprehension gain ($p < .05$; linear); r =.08 with factual gain (not significant).

In sum, the two things that stand out in the study by Solomon *et al.* are that certain qualities are related to achievement, and that achievement needs to be considered in a differentiated, rather than in a global manner.

There is a small set of related studies which report investigations of "explaining ability". This topic or characteristic is more microscopic than the dependent variables in the researches of Furst and of Solomon *et al.* Rosenshine's study will be presented here, in particular.

The data for this set of studies come from a single source, and are a peculiarly small sample of teaching behavior--although there is the virtue of standardization of "what" was taught. Forty-three teachers in high school social studies taught two 15-minute lessons on contemporary events in Yugoslavia and Thailand. Inter-class ability differences were adjusted via students' performance on a test following presentation of 15 minutes of recorded teaching on a third country. Relevant tests were administered following the Yugoslavia and Thailand lessons.

Rosenshine (reported in Nuthall, 1970, pp.25-26) investigated a large number of measures of teacher behavior, in relation to two sub-groups of teachers--the most and least effective, as measured by their classes' performances. The three variables that were significantly related to effectiveness (class performance, adjusted for level on the 'test" lesson) follow:

Gesture and Movement: the more able teachers showed a greater tendency to move around the room and gesture with hands, head, and trunk;

Rule and Example Patterns: the more able teachers showed a greater tendency to state rules both before and after discussing examples, while less able [successful] teachers tended to state the rule only once; either before or after the example.

Explaining Links: the more able teachers tended to make greater use of linking words such as "because, therefore, in order to, consequently, by means of, since, etc." (Nuthall, 1970, p.26).

From the various studies reported, it seems warranted to draw a few conclusions or generalizations. Since Rosenshine and Furst (1973) list 9 components that they found correlated with student achievement in the studies they reviewed, the conclusions from this overview will be compared with that list.

The Furst study indicated, for instance, the importance of (a) "extended" indirect teacher talk, (b) more positive than negative feedback to student talK, (c) more talk by students, (d) moderate "verbal structuring moves" and (e) moderate rates of questions and answers. Many of these aspects are similar to the components of the "indirect" composite category of Flanders' system of observation, which was found in this Furst study to be positively related to achievement. Wolfson's and La Shier's studies supported the significance of these kinds of behaviors also, indexed by the I/D ratio. They, and like behaviors and classroom interaction patterns, were also found important in Solomon _et al._'s study: Moderate "permissiveness vs. control," "energy vs. lethargy," and flamboyance vs. dryness." Lastly, Rosenshine's identification of physical "gesture and movement" may (or may not) betoken correlated psychological and interpersonal flexibility, in contrast to rigidity.

The descriptive names Rosenshine and Furst used for variables related to the above are "Variability," "Enthusiasm," "Criticism" (typically, a negative relationship to student achievement found), "Teacher Indirectness," and "Use of structuring comments" (1973, pp.156-7).

The second theme that can be found in the _conditions_ of student achievement, from the above studies, is much more specifically "cognitive," informational, or rational/mental in nature.

Furst cites, as promotive of student learning, teachers exhibiting a variety of logical processes, and less "lecturing." Rosenshine's comments on Furst's study added that there were more incidents of defining and interpreting in the high-achieving classes, compared to fact-stating. Solomon et al's study indicated "clarity" was important for gaining factual knowledge, but it was not related to comprehension. Then, Rosenshine's study of micro-units of teaching behavior indicated that both explaining links ("therefore," "since," etc.) and rule-example-rule patterns were important.

Rosenshine and Furst, in their review of the correlational studies (1973, pp.156-7) give "clarity" and "multiple levels of questions or cognitive discouse" as the categories that are related to the fore-going. --Thus, "task-oriented and/or businesslike" is the only one of their nine variables that is not paralleled in the above set. Hence, the general parallels between this review and the summary list from Rosenshine and Furst support the implicit argument of this review -- that these types of variables are the ones of apparent importance and therefore should take precedence.

The foregoing studies suggest companion circumstances as being important. First, interpersonally responsive interactions between teacher and students (and students and students), such as "moderate" control/ permissiveness, moderates rates of questioning and answering, and of lecturing, "more extended" student participation (rather than "rapid-fire" exchanges), etc. have been shown to be significant. These seem to be highly compatible with the indirect aspects of the Flanders system of classifying teacher influence. (See Table 1 of this appendix.)

The indirect influences would reasonably appear to incorporate at least two dimensions of learning environments: (1) an atmosphere of warmth, more egalitarianism (though not pupil control), interpersonal respect; these are the value aims of Flanders, in fact. (2) more opportunity for student participation, for instance, in terms of some initiative in asking questions, and practice (overtly and covertly) in using ideas, thinking logically. Consequently, the Flanders classification system is recommended for inclusion in studies of physics courses or any college course.

The second generalization from the studies and reviews is the importance of cognitive components of teaching and learning. Variety of logical processes, more instances of teachers' defining and interpreting vs. fact-stating, the utility of "explaining links" and rule-example-rule patterns are indicators of this broad class of phenomena. The system that has been specifically designed to differentiate this domain is the Bloom taxonomy of cognitive behavior(1956). Its basic premise is that there are levels of cognitive activities that differ in complexity and that can be arranged in a hierarchical order with respect to complexity and inclusiveness. The simpler levels are factual, then "translation," moving into interpretation, applications of knowledge, analysis, to the most complex, namely synthesis, and evaluation. Brown and his co-workers at Florida (Brown et al, 1971 and Webb, 1970) have refined and further systematized Bloom's work, for readier application to teaching events. Thus, this Florida system is recommended here. (See Table 2 of this appendix).

Additionally, a serious attempt should be made to examine and classify the cognitive levels and their relative weightings which are minimally necessary to meet the achievement performance criteria of the course, most centrally the examinations. Not only has research shown that different conditions of teaching/learning are related to different types of criterion performance (cf. Solomon et al.), but it cannot be taken for granted that differentiated classroom cognitive tasks, in the Bloom sense, are appropriately reflected in the things the students "really" have to learn, namely what is required of them in order to earn their grades.

Rosenshine and Furst point out (1973, p. 158) that, where tested, such findings as have been presented here, based on correlational studies, have not been replicated in experimental studies, i.e., those using explicit manipulation of teaching behaviors as independent variables. One view of this non-replication would be that the variables manipulated are not the critical ones . A different view would be that something(s) critical is changed, but unmeasured -- in both types of studies -- such that an experimentally designed course is not comparable to the 'natural' ones which have been observed. Although a content analysis of this physics course would constitute a correlational study, studies of any college courses are too scarce. Therefore it seems worth the effort to undertake a study of this physics course with respect to the cognitive and interactive variables recommended. A much more ambitious, presumably experimental, program of studies would be required to investigate the sources of the discrepancies in findings of the two types of studies.

BEST COPY AVAILABLE

Table 1

Summary of Categories for the Flanders System of

Interaction Analysis in its Regular Ten-Category Form*

Teacher Talk

Indirect Influence

1. Accepts Feeling: accepts and clarifies the feeling tone of the students in a non-threatening manner. Feelings may be positive or negative. Predicting or recalling feelings are included.

2. Praises or Encourages: praises or encourages student action or behavior. Jokes that release tension, not at the expense of another individual, nodding head or saying "um hm" or "go on" are included.

3. Accepts or Uses Ideas of Student: clarifying, building, or developing ideas or suggestions by a student. As teacher brings more of his ideas into play, shift to category five.

4. Asks Question: asking a question about content or procedure with the intent that a student answer.

Direct Influence

5. Lecturing: giving facts or opinions about content or procedure; expressing his own ideas, asking rhetorical questions.

6. Giving Direction: directions, commands, or orders to which a student is expected to comply.

7. Criticizing or Justifying Authority: statements intended to change student behavior from non-accept[able] to acceptable pattern; bawling someone out; stating why the teacher is doing; extreme self-reference.

Student Talk

8. Student Talk--Response: talk by students in response to teacher. Teacher initiates the contact or solicits student statement.

9. Student Talk--Initiation: talk by students which they initiate. If "calling on" student is only to indicate who may talk next, observer must decide whether student wanted to talk. If he did, use this category.

[Neither of above]

10. Silence or Confusion: pauses, short periods of silence and periods of confusion in which communication cannot be understood by the observer.

* From Flanders, N.A., Teacher Influence, Pupil Attitudes, and Achievement. Coop. Res. Monogr. No. 12, OE 25040. Washington: U.S. Dept. of H.E.W., 1965. [In turn, taken from p. 35 of Ober, R.L., The reciprocal category system, J. Res. & Devel. in Educ., 1970, 4(1), 34-50.]

BEST COPY AVAILABLE

Table 2

The Florida Taxonomy of Cognitive Behavior*

1.10 Knowledge of Specifics

1. reads
2. spells
3. identifies something by name
4. defines meaning of term
5. gives a specific fact
6. tells about an event

1.20 Knowledge of Ways and Means of Dealing with Specifics

7. recognizes symbol
8. cites rule
9. gives chronological sequence
10. gives steps of process, describes method
11. cites trend
12. names classification system or standard
13. names what fits given system or standard

1.30 Knowledge of Universals and Abstractions

14. states generalized concept or idea
15. states a principle, law, theory
16. tells about organization or structure
17. recalls name of principle, law, theory

2.00 Translation

18. restates in own words
19. gives concrete example of an abstract idea
20. verbalizes from a graphic presentation
21. translation of verbalization into graphic form
22. translates figurative statements to literal statements or vice versa
23. translates foreign language to English or vice versa

3.00 Interpretation

24. gives reasons (tells why)
25. shows similarities, differences
26. summarizes or concludes from observation of evidence
27. shows cause and effect relationship
28. gives analogy, simile, metaphor
29. performs a directed task or process

4.00 Application

30. applies previous learning to a new situation
31. applies principle to new situation

(continued)

* From Brown, Bob Burton et al., The Florida Taxonomy of Cognitive Behavior, 1971.

Table 2 (cont'd)

32. applies abstract knowledge in a practical situation
33. identifies, selects, and carries out process

5.00 Analysis

34. distinguishes fact from opinion
35. distinguishes fact from hypothesis
36. distinguishes conclusions from statements which support it
37. points out unstated assumption
38. shows interaction or relationship of elements
39. points out particulars to justify conclusion
40. checks hypothesis with given information
41. distinguishes relevant from irrelevant statements
42. detects error in thinking
43. infers purpose, point of view, thoughts and feelings
44. recognizes bias or propaganda

6.00 Synthesis

45. reorganizes ideas, materials, processes
46. produces a unique communication or divergent idea
47. produces a plan, proposed set of operations
48. designs an apparatus
49. designs a structure
50. devises a scheme for classifying information
51. formulates hypothesis, intelligent guess
52. makes deductions from abstract symbols, propositions
53. draws inductive generalization from specifics

7.00 Evaluation

54. evaluates something from evidence
55. evaluates something from criteria

References for Appendix F

Bloom, B. S. (ed.), Taxonomy of educational objectives, the classification of educational goals, handbook I: Cognitive domain. New York: David McKay, 1956.

Brown, B. B., Ober, R. L., Soar, R. S., & Webb, J. N., The Florida Taxonomy of Cognitive Behavior. Gainesville, Fla.: Institute for the Development of Human Resources, College of Education, Univ. of Florida, 1971 (mimeo).

Flanders, N. A., Teacher Influence, Pupil Attitudes, and Achievement. Coop. Res. Monograph No. 12, OE 25040. Washington: H.E.W., 1965.

Nuthall, G. A., A review of some selected recent studies of classroom interaction and teaching behavior. AERA Monograph Series on Curriculum Evaluation, No. 6, Classroom Observation, 1970.

Rosenshine, B. and Furst, N., The use of direct observation to study teaching. In Travers, R. M. W. (ed.), Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973.

Rosenshine, B., Teaching behavior related to pupil achievement: a review of research. In Westbury, I. and Bellack, A. A. (eds.), Research into Classroom Processes. New York: Teachers College, 1971.

Rosenshine, B., Some criteria for evaluating category systems: an application to the topic classification system. AERA Monograph Series on Curriculum Evaluation, No. 6, Classroom Observation, 1970.

Webb, J. N., Taxonomy of cognitive behavior: a system for the analysis of intellectual processes. J. Res. & Devel. in Educ., 1970, 4(1), 23-33.

Wolfson, M. L., A consideration of direct and indirect teaching styles with respect to achievement and retention of learning in science classes. J. Res. in Science Teach., 1973, 10(4), 285-290.