ABSTRACT
        Biology classes of 15 teachers in three high schools
were provided with a response sheet and requested to rate each
ten-minute segment of lessons on each of three contrasts: too
fast/too slow; interested/bored; and understand/don't understand.
They were also requested to make specific comments if they felt them
necessary. It was found that students were able to give effective
feedback to teachers, and that this feedback caused changes in
teacher behavior. Analyses of the data were most meaningful when the
changes within a single class were considered: changes in different
directions in different classes obscured significant effects. Tabular
and graphical data presentations are appended. (AL)

A STUDY OF THE EFFECT OF TIMED PUPIL FEEDBACK ON THE TEACHING BEHAVIORS OF BIOLOGICAL SCIENCE TEACHERS

NATIONAL ASSOCIATION FOR RESEARCH IN SCIENCE TEACHING

April 4, 1972
Chicago, Illinois

John H. Novak
Gene W. Moser

Science Education
2609 Cathedral of Learning
University of Pittsburgh
Pittsburgh, Pennsylvania 15213

TABLE OF CONTENTS

Appendices

I. Introduction

An inability to perceive and react to student impressions
(impressions such as: too fast, too slow, boredom, interest,
understanding, etc.) hinders a teacher's ability to encourage
student participation, maintain interest, and promote the
inquiry strategies suggested in many of the new curricula.
If teachers are to develop the ability to assess and utilize
students' reactions to their classes, they must recognize pupils
as integral parts of the instructional process. It is the
purpose of this study to determine if the use of a pupil reaction
device will cause teachers to become more aware of the attitudes
of their pupils. This awareness will be reflected in the teacher's
willingness to vary teaching behavior in order to promote more
positive student feedback.

II. Related Literature

Grace E. Bird, in 1917, tried to describe an effective
teacher by using feedback given by pupils.[1] The children were
asked to write a description of their favorite teacher and to
tell why the teacher was their favorite. What emerged was that
students favored teachers who were responsive to their immed-
iate needs.

Roy C. Bryan studied the use of pupil ratings to improve
teacher effectiveness in 1938.[2] The findings of the study, as
reported in School Review, indicated teachers did react to
student ratings given them at the beginning of the school year
as evidenced by an increase in favorable scores given at the
end of the year. Albert, observing that teachers are usually

3

only rated by a principal or supervisor, developed his own
rating scale to be used by students to determine their likes
and dislikes in teacher behavior.[3] He found that student
ratings did have an effect on teacher performance. Earl C.
Bowman, reporting on pupil ratings of student teachers, used
ratings of student teachers prepared by pupils and compared
them with ratings given the same student teachers by critic
teachers.[4] He found very little agreement between pupil and
critic teacher judgment as to the presence or absence of desir-
able traits described by the Purdue Rating Scale of Teacher
Efficiency.

Remmers has pointed out that student evaluation has proven
"reliable", "convenient", "useful", and "valid".[5] One of the
more interesting major generalizations drawn from research using
the Purdue Scale is: Ratings of teachers by groups of 25 or
more students are as reliable as any other means of rating.
Smith reported on the Student Evaluation of Teachers Committee,
where the "evaluator" asked the pupils, as a group, pertinent
subjective questions about the teacher.[6] Smith found the pupils
to be". . . competent judges of teaching skill."[6] Hedges and
MacDougall urged that a monitoring device be implemented to
tell the teacher what the student perceives to be going on in
a classroom so that he may have immediate feedback of pupil
reaction.[7] William D. Coats, using the Teachers Image Question-
naire as prepared by Educator Feedback Center at Western Mich-
igan University, found the single most important item which
appeared to influence a teacher's rating by pupils was whether
or not the students liked the teacher.[8]

III. Statement of the Problem

This study attempted to show that a teacher's pupils can
be used to provide an effective and sensitive measure of teach-
ing behavior in the form of timed hard copy feedback given the
teacher at the end of each lesson. This study further attempts
to describe the behavioral changes brought about by the pupils'
hard copy feedback and the permanancy of any change once the
pupils' written feedback stops.

IV. Procedure

A population of 15 tenth grade biological science teachers
were chosen from three separate high schools within Allegheny
County during the 1971-72 school year. For each of the teachers
and their pupils taking part in the study, the following pre-
liminary data were gathered.

Pupils	Teachers
(1) age	(1) age
(2) grade level	(2) school where presently teaching
(3) I.Q. -	(3) school graduated from
(4) socio-economic group	(4) sex
(5) school attending	(5) years experience teaching
	(6) years experience teaching subject
	(7) number of credit hours in teaching field

Nine of the total 15 teachers chosen for the first three
months of the study were designated as Group I. Audio tapes
of one class were made of each of the nine teachers. The
class period chosen for each teacher was the only class for that
teacher to be used throughout the three months. Two audio
tapes per week were taken of the experimental class. The time
of the taping was chosen at random. (The reasons for taping

the classes in month (1) was to establish a base line of teaching behavior.)

During the second month, in addition to continued audio taping, the cooperating teacher was asked to pass out to his pupils the rating sheet (see Appendix A) with these instructions: "I want you to help me to help you learn biology. The sheet of paper you have on the desk in front of you is a means for you to tell me how I did today in trying to teach you biology. You will notice that the blocks from left to right are marked with comments and at the end of the blocks is a line with numbers ranging from 10 to 50. What you are to do is, if at the end of ten minutes you think I was going too fast, you check "too fast". If I was boring you, check "boring". If it was interesting, check "interesting", etc. If you wish to say something else, just write it on the line provided. Then the next time you check will be at the end of the next ten minutes (or 20 minutes after the lesson started), continuing in ten minute intervals all the way to the end of the lesson. If you forget or are too busy to check the clock, just skip over it and check when you have time, but do it in the appropriate time line. For example, it is now 12:00. If the next time you think to check the sheet it is 12:20, check the 20 minute line, not the 12:10 line." The teacher continued to explain these directions until he was satisfied that he was understood. He then gave the planned lesson for the day and collected the sheets at the end of the period. If he found they did not follow instructions, he went over the instructions again the next day.

The teacher was asked to use the rating sheets at least

twice per week during the first three weeks of the second month and only if he or his students felt they needed them during the last week.

The teacher was asked to review the feedback written by his pupils on the sheets given him at the end of the experimental class to see how his students received his lesson. The cooperating teacher was also urged to respond to the sheets in such a manner as would emphasize behaviors and procedures that the pupils felt were good and to modify those behaviors or procedures that in the pupils' judgment needed changed.

During the third month, the nine teachers were asked not to use the pupil feedback form. Audio taping of the classes continued however, and three audio tapes were chosen for analysis.

The following outline describes the schedule and the use of the response instrument.

Outline of Procedure

Month (1)

1. Nine teachers chosen.(Group I)

2. Background data taken.

3. Audio tapes taken of each teacher during designated class period(s).

4. Three audio tapes chosen at random for each teacher.

Month (2)

1. Group I teachers explain rating device to pupils.

2. Teachers utilize rating device.

    a. First week - two times

    b. Second week - two times

       c. Third week - two times

       d. Fourth week - as the teacher feels necessary.

   3. Audio tapes taken of selected classes.

   4. Three audio tapes chosen at random for each teacher.

Month (3)

   1. No rating device used by Group I teachers.

   2. Audio tapes taken and three chosen at random for each teacher.

Month (4)

   1. Six teachers chosen. (Group II)

   2. Background data taken.

   3. Audio tapes taken of each teacher during designated class period(s).

   4. Three audio tapes chosen at random for each teacher.

Month (5)

   1. Group II teachers explain rating device to pupils.

   2. Teachers utilize rating device.

       a. First week - two times

       b. Second week - two times

       c. Third week - two times

       d. Fourth week - as the teacher feels necessary.

   3. Audio tapes taken of selected classes.

   4. Three audio-tapes chosen at random for each teacher.

Month (6)

   1. No rating device used by Group II.

   2. Audio tapes taken and three chosen at random for each teacher.

V. Findings

Preliminary analysis

Three separate high schools participated in the study. School A is an urban, middle class high school with an enrollment of 3,200. High School B is an urban high school with an enrollment of approximately 500 lower socio-economic students. High School C is in an upper-middle class suburb with an enrollment of over 2,000.

Facilities at each school, provided for biology, ranged from a room with only a teachers' desk and student chairs, to elaborate laboratory set-ups with a green house and animal rooms. Only two of the 15 groups were scheduled for biology for more than five periods per week. These two had two extra laboratory periods per week: for a total of seven. The average class size was 32 students with a range of 27 - 36. The mean I.Q. of all participating students was 104 with a range of class mean I.Q.'s of 93 - 121. Pupils ages averaged 15.4 years and 90 per cent of the students were in the tenth grade.

Teachers participating in the study ranged in age from 23 - 50 years old and had experience in teaching biology which ranged from 1 - 15 years. The mean years experience was four years. There were 12 men and three women in the group. They graduated from ten different colleges and universities.

Three tapes for each teacher for each month of the study were selected at random and coded using a modified Parakh Interaction Analysis System.[9] Each selected tape was further coded using a Gallagher-Aschner Questioning Category System Analysis.[10] The monogram codes of the Parakh were then placed

in a six-set matrix (Appendix B) and the per cent of lecture,
per cent of discussion plus inquiry, and the per cent of tran-
sition* was calculated for each six-set for each teacher.[11]

Data Analysis

The entire group of teachers were checked for significant
change over the three months with respect to the per cent of
lecture discussion and transition, using the chi square test

01 for k independent samples.[12] The results showed: __No signif-__
__icant difference in the six-set analysis of the group teacher__
__behavior over the three months period.__

Group teacher behavior was again inspected comparing
months one to two, months one to three, and months two to three.

02 The results indicated: __No significant differences in the six-__
__set analysis of the group teacher behavior was found between__
__any pairs of months.__ This finding will be discussed below.

Data for each teacher were then tested for significant
differences in the six-set analysis of behavior over the entire
three months, and between months one and two, one and three,

03 and two and three. __Significant differences in six-set analysis__
__were shown for most teachers.__ (See Appendix E Table 1)

The first five sets of completed pupil reaction forms
collected by all 15 teachers were totalled and treated for
per cents (See Table 2, Appendix E and Graphs 1 - 4, Appendix
C). An interpretation of Graphs 1 - 4, Appendix C showed:

04 __Little difference in the per cent of reaction in any category__
__over the entire time frame.__ It is then reasonable to assume
that the pupils were not reacting in a "set pattern" regardless

---

*Transition is a term to describe modes of performance which
lead into or out of the major modes: see Appendix B.

05 of teacher behavior. However, an increase in the total entries in the 20 - 30 minute time block may be indicative of the time-honored tradition that nothing much happens in a class at the beginning and at the end.

The data of four teachers were selected at random. Each teacher"s pupil reaction sheets were compared from response set one through set five for significant change in the per cent 06 of each category checked by the pupils. Significant differences were found in the responses from set one through set five.

It was reasoned that if the student-rating forms were having a positive effect on most of the teacher's behavior, then the ratio of all the students' positive to negative reactions 07 would increase as time went on. It was found that the direction of change was positive for responses made by the students. (See Graphs 1 - 3, Appendix D)

Product-moment coefficients of correlation and regression analysis[13] treatments were used to test the significance of relationships between student responses and the six-set analysis of teacher behaviors for the total lessons coded during the 08 treatment month. Only five per cent of the test values were found to be significant at the five per cent level. The lack of significance was found to be attributable to the smoothing of data when grouped for lessons averaging 42 minutes in length. It is quite logical that when considering the approximately 30 students who respond to behavior in a classroom every ten minutes, the total behavioral responses cannot reliably define the environment.

It was then decided to select one 10-minute time interval

from each lesson and test the relationship of the pupil reactions to the six-set analysis of that interval. The lessons were chosen on the basis that they would be the first lesson taught by the teacher, and which had a corresponding set of pupil-rating sheets. The 20 - 30 minute time interval was chosen because it had an overall high N of responses given by pupils. Product-moment coefficients of correlation and regression analyses were then used to test the significance of relationships between student responses and six-set analysis of teacher behaviors for 14 of the teachers during the given 10 minute

09 time frame. It was found that three of the nine tests showed significant correlations at the five per cent level and two at the ten per cent level. (See Table 3, Appendix E).

Analysis of Findings

The ability of students to sensitively and effectively give hard copy feedback to teachers is pointed out by findings 01 and 02, where it was shown there were no significant differences between the three months of group teacher behavior. This can be expected if individual differences of teachers are considered by students and their reactions are to their teacher only and their reactions change as the teacher behavior changes. Although teacher A may change from a more lecture oriented behavior to a more discussion oriented behavior in response to the students' feedback, this change could be cancelled by teacher B's "movement" in the opposite direction in response to his students. These changes of behavior by different teachers is unique and the effect is that grouped data-treatments become less meaningful.

There seemed to be no pattern for filling out the reaction forms (Findings 04 and 05). The students indicated they were as bored the first ten minutes as they were during the last ten minutes. Of course it was equally probable for them to be bored in the intervening intervals. On the other hand, the number of students responding at any one time interval is highly dependent to how close the time is to the middle of the lesson. The Parakh Analysis[9] of the original tapes point out a possible explanation to this finding. There was little likelihood of anything of import happening in the first ten and last ten minutes of a class period. Most of this time was spent in processing routine procedures, generally wasteful activity, or conversation on the part of students and teacher.

If students were reacting to their teacher and reporting it accurately as they perceived it on the reaction form, then as the teacher behavior changed, so too should the reaction form responses change. Finding 06 points out that there were significant changes in the day to day rating of each individual teacher sampled.

The overall change in reaction for all pupils is illustrated by finding 07. The changes of ratios in "favorable" reaction by the students are indicative of teacher change. The change in teaching behavior is further illustrated by (Table 1 Appendix C) finding 03. With the exception of teacher C, all teachers changed significantly over the three month period. The major changes occured between the first and second months and first and third months. The lowest amount of change was between the second and third months, however, only four chi

square tests were found not to be significant below the .05 level. This seems to be due to the teachers changing to a lesser degree after the second month of the study.

The general direction of change in overall behavior during the three months, although not significant, was toward more lecture. During the treatment month, however, the direction was non-determinant for the group. Some teachers moved toward giving more lecture while others moved toward more discussion and transition. One of the most interesting findings was O9. A positive correlation exists between the amount of lecture and the percentage of students responding to "understand" on the rating sheet and a negative relationship to the amount of discussion and transition. This then could account for those teachers who maintained or increased their amount of lecture during and after the treatment month. Finding O9 also shows a negative correlation between the "good" response and the amount of lecture and a positive correlation between transition and "good".

It would seem, for these groups of teachers and students, the students feel that in order to "understand", lecturing is desirable and the amount of discussion with their peers seems to decrease their willingness to check the "understand" reaction square. On the other hand, teacher led discussion, described as transition, is considered "good" by the students.

The reasoning behind this seeming paradox could be any or all of the following:

1. If a teacher leads a discussion or "causes" one,
   students seem to regard his performance as of more

14

note than their peers,

2. It is a good topic, but need not necessarily be under-
   stood,

3. If one checks "good" they may not find it necessary
   to check any other square,

4. A lecture need not necessarily be "good" but it may
   be understood;

5. There seems to be some discriminators operating for
   topic differences.

VI. Conclusions

There are two major conclusions to be drawn from the
findings in the previous section. First, students do give
meaningful, timed, hard copy feedback responses to teachers'
behavior. Second, that if this feedback is even minimally
used, as it was in this study, it can be expected to cause
teachers to change their behavior and for many maintain this
behavior, as evidenced by both six-set analysis and student-
rating analysis. These changes in behavior are not regarded
as any improvement in performances, nor do they indicate any
quality in teaching.

It is felt that such rating devices, employed on a day-to
day basis, or whatever time schedule can be organized, should
be used by supervisors who wish to help their present teachers
and by those individuals responsible for the training of future
teachers.

The data further indicates that, in order to look at any
affects and any relationships between student feedback and
teacher behavior, it would be more valuable to look at small

sections of time and to focus on those small areas which constitute a one teacher one class interaction.

Appendix A
Sample Student Reaction Sheet

Teacher_____   Date_____

Too Fast	Interested	Understand	Good	12:00
Too Slow	Bored	Don't Understand		Comment

Too Fast	Interested	Understand	Good	12:05
Too Slow	Bored	Don't Understand		Comment

Too Fast	Interested	Understand	Good	12:10
Too Slow	Bored	Don't Understand		Comment

Too Fast	Interested	Understand	Good	12:15
Too Slow	Bored	Don't Understand		Comment

Too Fast	Interested	Understand	Good	12:20
Too Slow	Bored	Don't Understand		Comment

Too Fast	Interested	Understand	Good	12:25
Too Slow	Bored	Don't Understand		Comment

Too Fast	Interested	Understand	Good	12:35
Too Slow	Bored	Don't Understand		Comment

Appendix B

Six Set Matrix

	6:0	5:1	4:2	3:3	2:4	1:5	0:6
6:0							
5:1							
4:2							
3:3							
2:4							
1:5							
0:6							

Cumulative Six Set

Lec- ture	Transition	Trans- ition
Trans- ition	Discussion	Trans- ition
Trans- ition	Transition	Trans- ition

Appendix C   Graph 1

Contrast of the Per Cent of Total Student
Responses to Too Fast or Too Slow for Each Time Interval of the Lesso

Per Cent
of
Responses

.30

.20

.10

C   10 min.   20min.   30min.   40min.   50m;

Minutes into Lesson
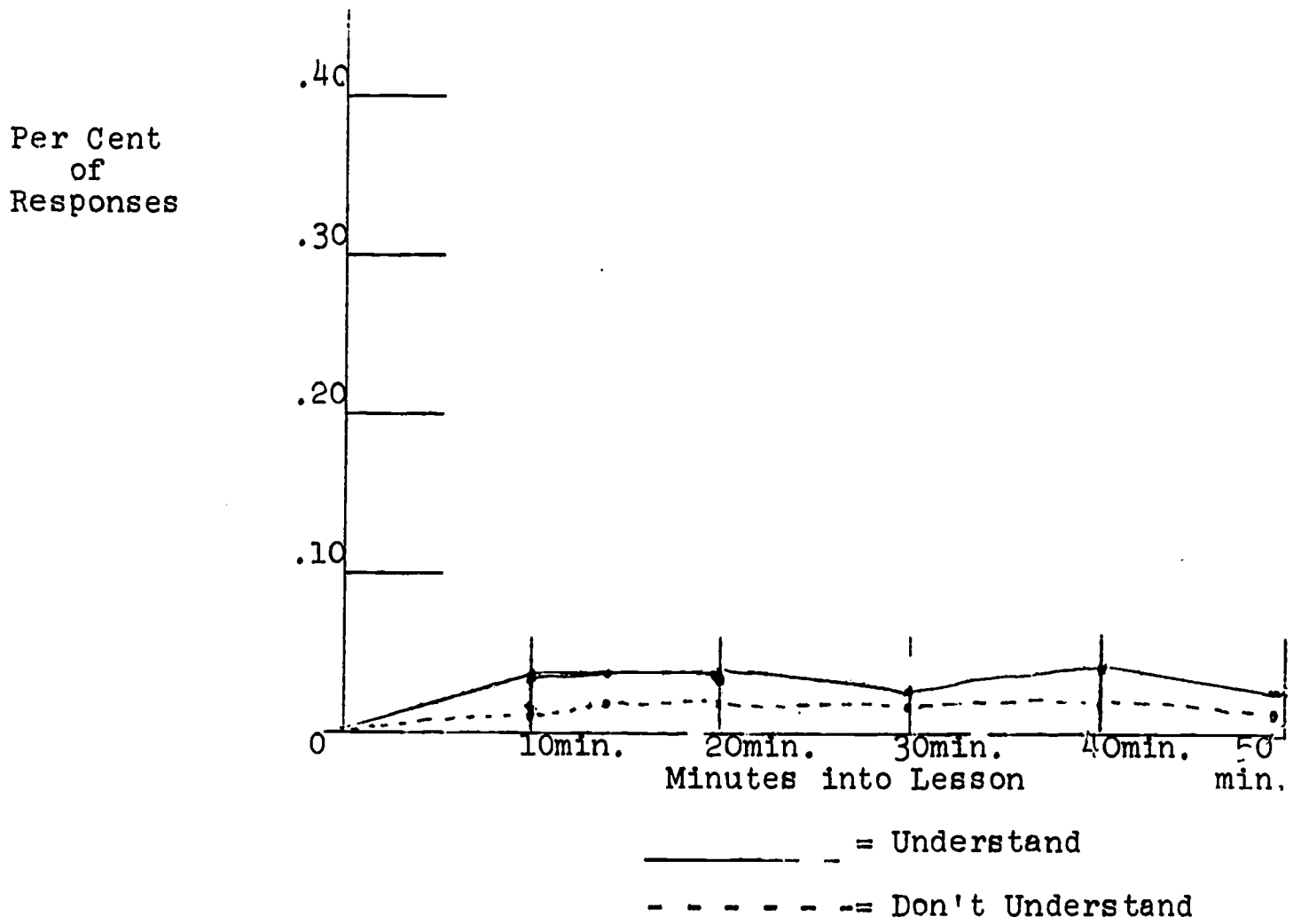
———————— = Too Fast

- .. .. - - - - -= Too Slow
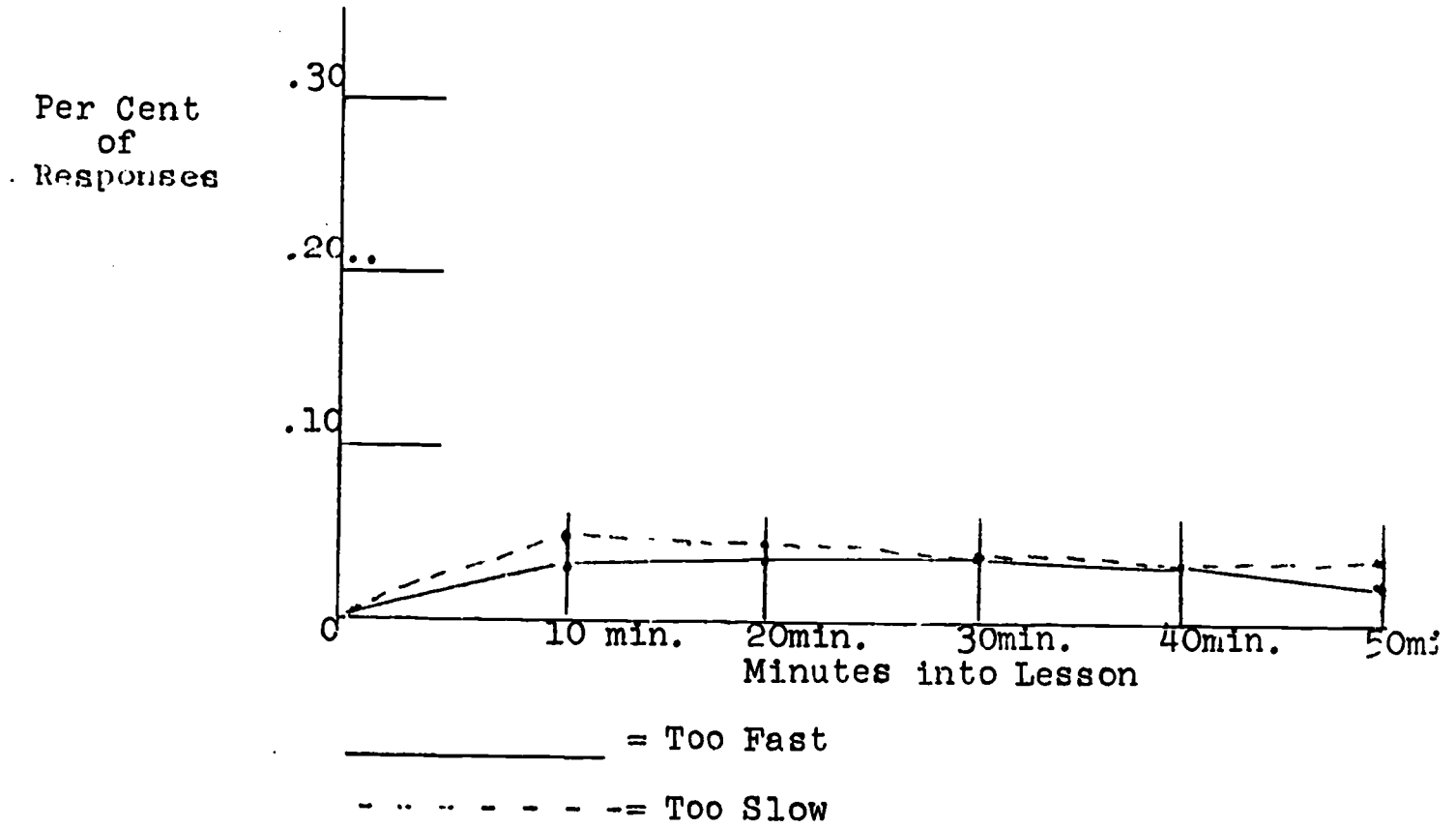
Appendix C   Graph 2

Contrast of the Per Cent of Total Student Responses
to Interested or Bored for Each Time Interval of the Lesson

Per Cent
of
Responses

```
  .40

  .30

  .20

  .10

   O      10min.   20min.   30min.   40min.   50mi.
                 Minutes into Lesson
```

_____ = Interested

- - - - - - = Bored

Appendix C Graph 3
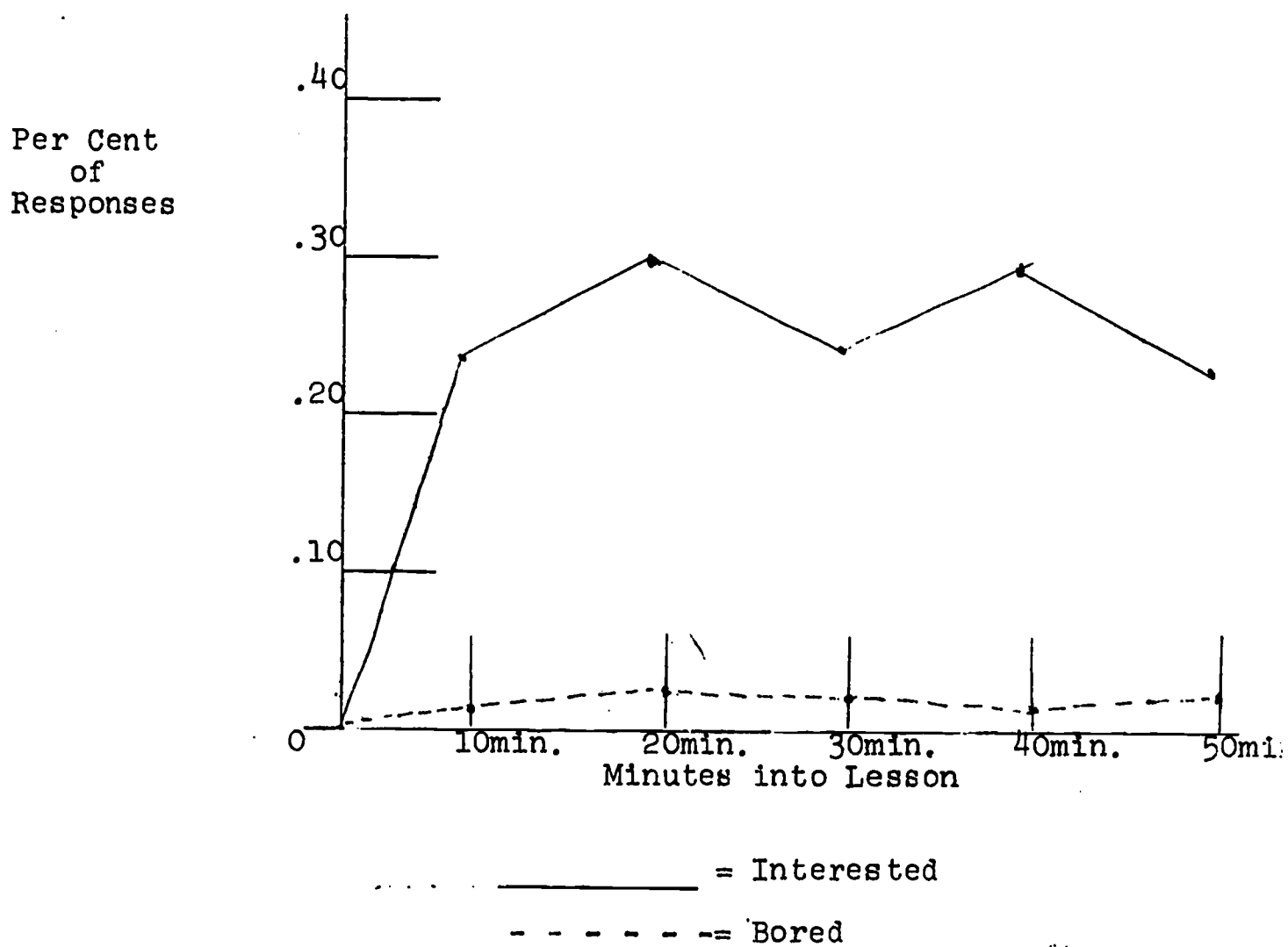
Contrast of the Per Cent of Total Student Responses
to Understand or Don't Understand for Each Time Interval of Lesson

Per Cent
of
Responses

_____ _ = Understand

- - - - - - -= Don't Understand

Appendix C   Graph 4

Per Cent of Total Student Responses
to Good for Each Time Interval of the Lesson

Per Cent
of
Responses

.40

.30

.20

.10

0

10min.   20min.   30min.   40min.   50 m

Minutes into Lesson
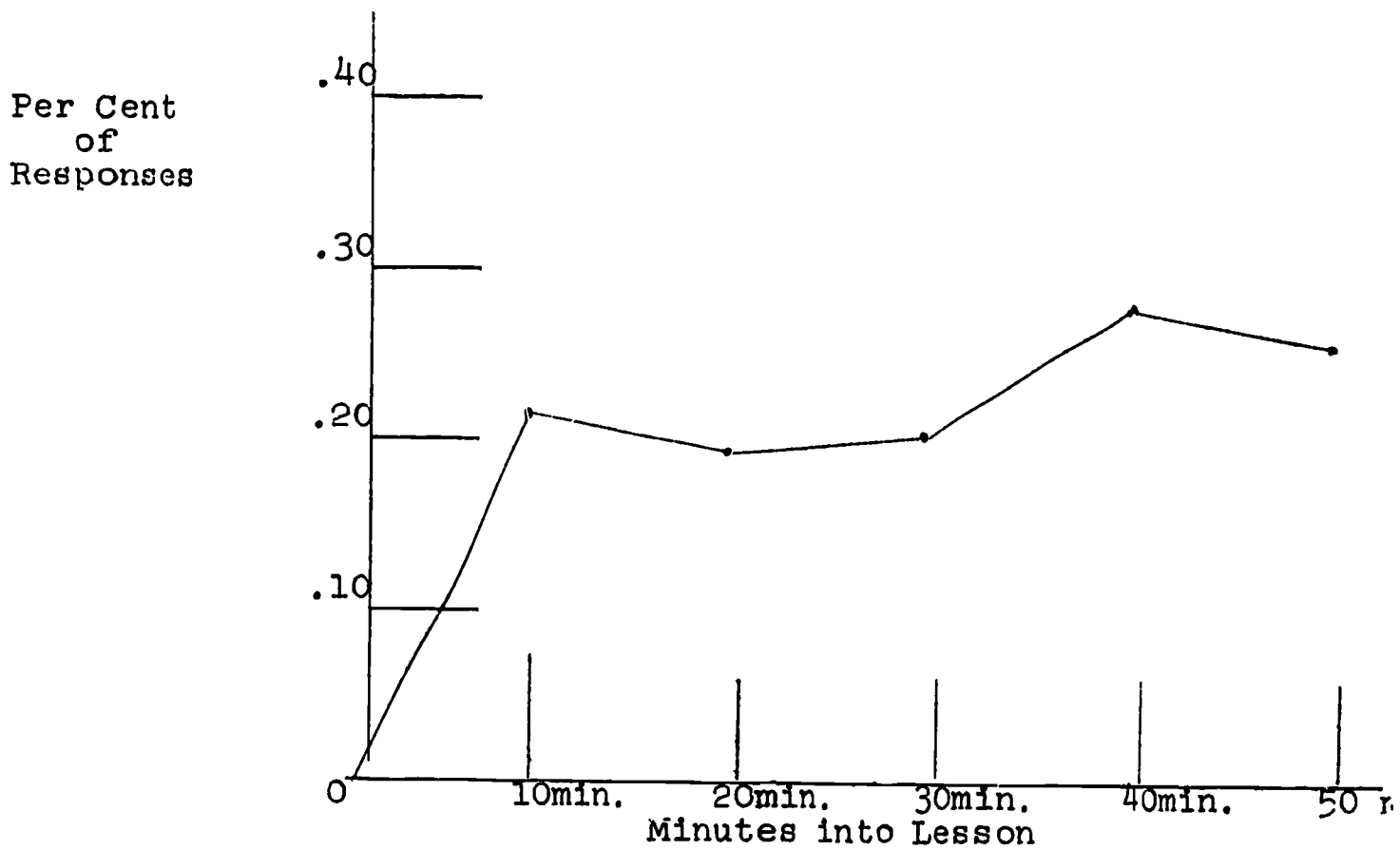
Appendix D   Graph 1

Overall Ratio of Positive/Negative Student
Reations for All Teachers During the Second Month of the Study
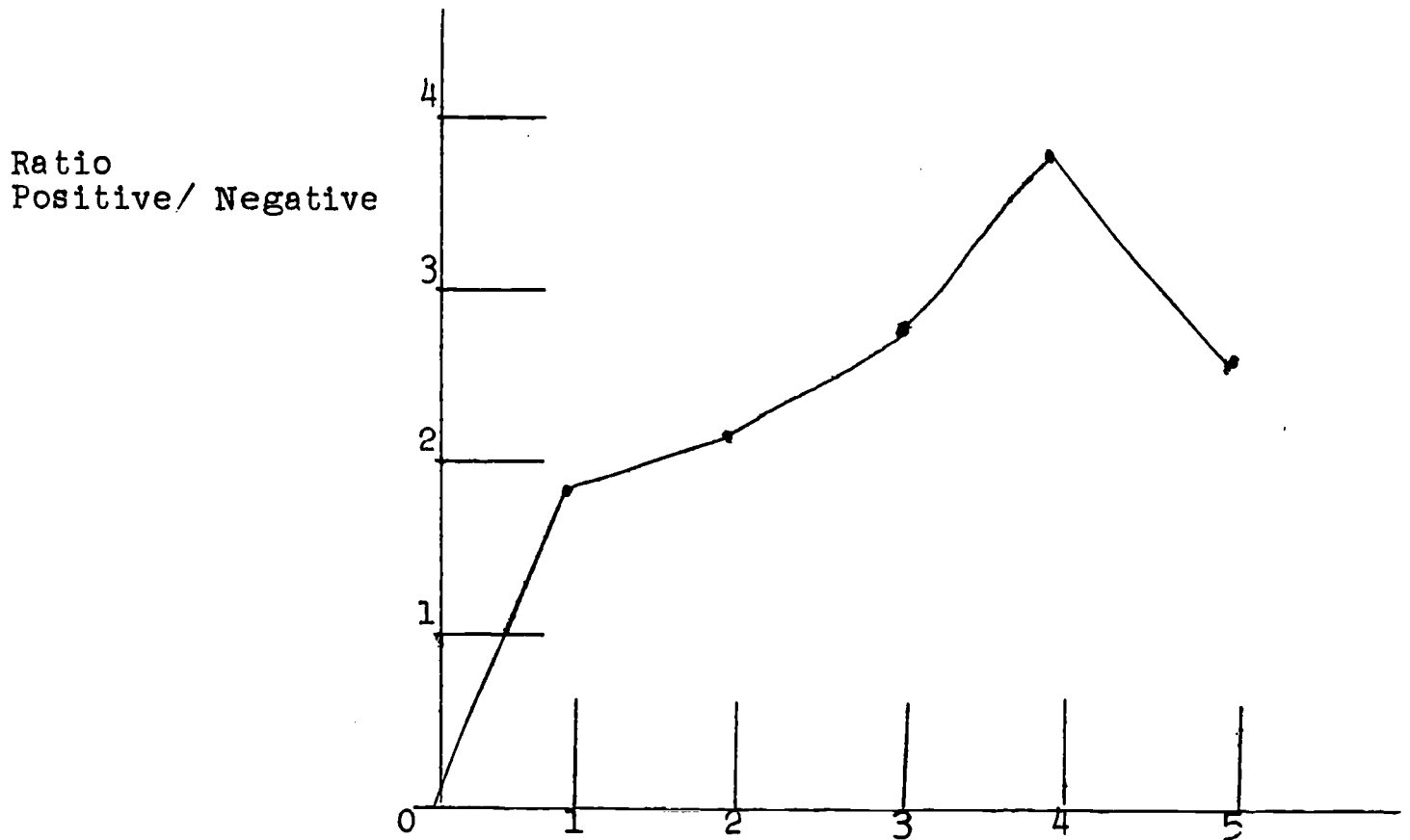
Ratio
Positive/ Negative

1 = First time sheet used
2 + Second time sheet used
3 = Third time sheet used
4 = Fourth time sheet used
5 = Fifth time sheet used

Appendix D   Graph 2
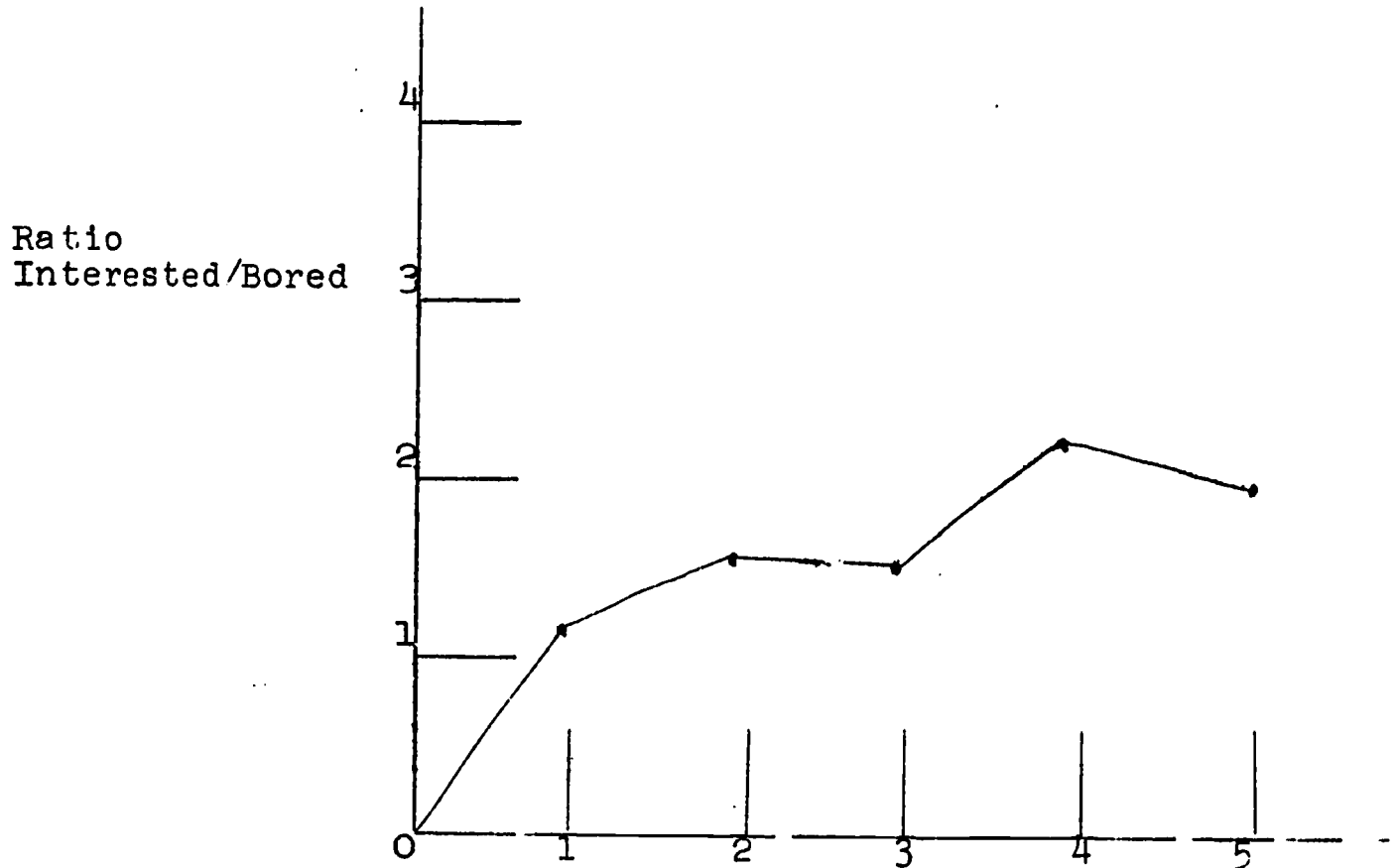
Overall Ratio of Interested/Bored Student
Reactions for All Teachers During the Second Month of the Study

1 = First time used
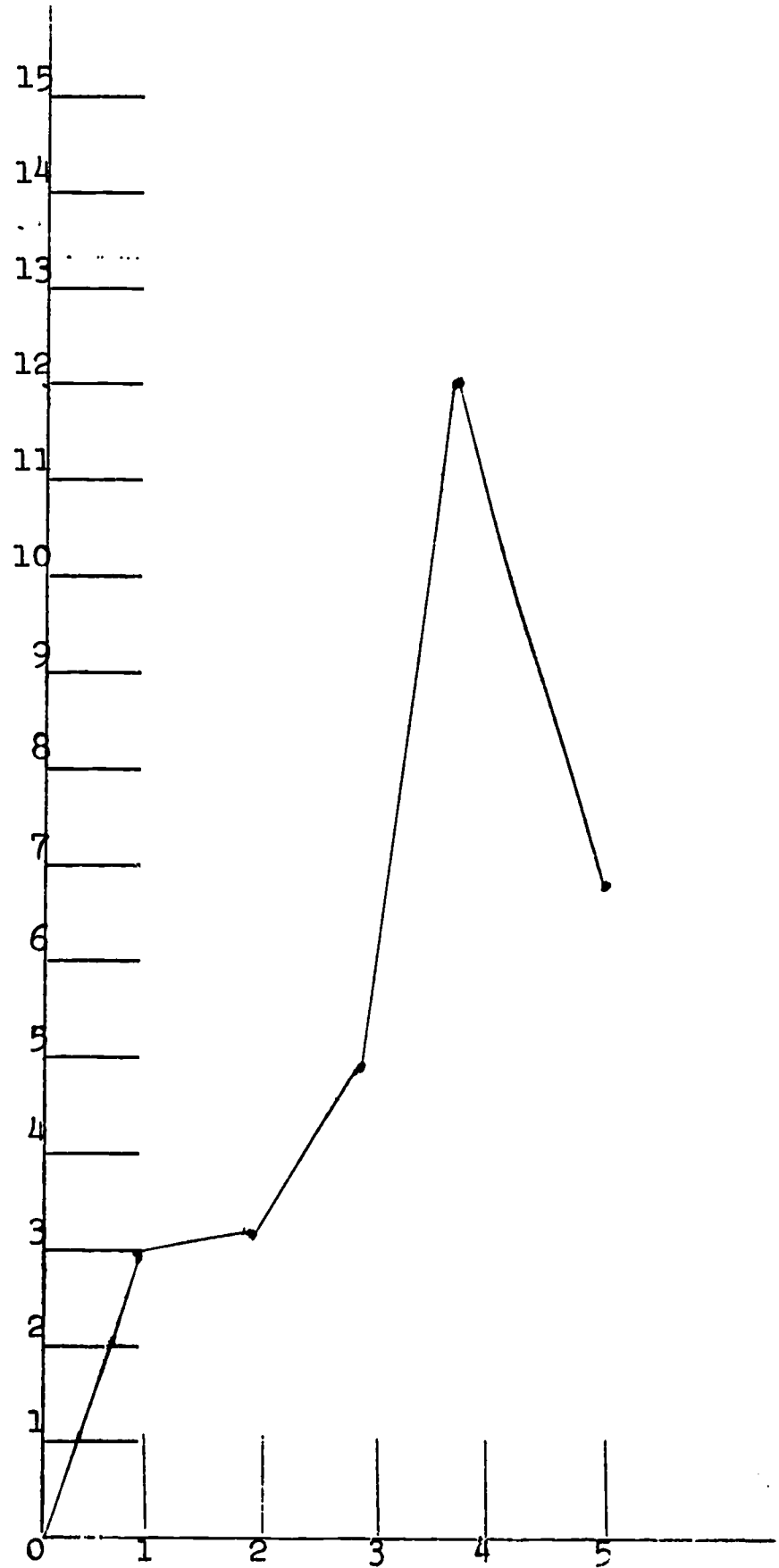2 = Second time used
3 = Third time used
4 = Fourth time used
5 = Fifth time used

Appendix D Graph 3

Overall Ratio of Understand/Don't Understand
Student Reactions for All Teachers During the Second Month of the Stu

Ratio
Understand/Don't Understand

1 = First time sheet used
2 = Second time sheet used
3 = Third time sheet used
4 = Fourth time sheet used
5 = Fifth time sheet used

Chi Square Values of each Teacher
and their Accompanying Levels of Significance
Comparing All Three Months
Month 1 to Month 2, Month 1 to Month 3, and Month 2 to Month 3

Teacher	All Months Chi Square	Level of Signifi- cance All Months	Month 1 to 2 Chi Square	Level of Signif- icance Months 1 to 2	Months 1 to 3 Chi Square	Level of Signif- icance Months 1 to 3	Months 2 to 3 Chi Square	Level of Signif- icance Months 2 to 3
A	18.4^a	.001	47.56	.001	40.21	.001	19.9	.001
B	78.38		38.48	.01	42.69	.001	10.65	.05
C	4.33	.50	10.27	.15	69.31	.001	16.81	.015
D	55.33	.001	36.36	.001	19.13	.001	17.28	.002
E	23.08	.001	12.57	.02	33.44	.001	6.04	.15
F	25.87	.001	39.92	.001	7.73	.10	33.12	.001
G	71.90	.001	5.97	.25	12.43	.02	9.48	.05
H	24.09	.001	20.00	.001	.03	.50	18.31	.001
I	99.52	.001	30.61	.001	29.20	.001	20.82	.001
J	23.19	.001	4.15	.50	5.48	.25	5.82	.25
K	13.36	.01	23.17	.001	20.88	.001	9.68	.05
L	12.56^b	.02	36.98	.001	5.42	.05	4.82	.10
M	19.12	.001	27.71	.001	7.72	.10	5.07	.25
N	41.93	.001	19.44	.001	23.98	.001	26.88	.001
O	12.02	.02	2.44	.50	6.81	.001	13.04	.001

a. To be read as 9.488 at .05% level of significance for 2 degrees freedom.
b. Exception to above to be read as 5.991 at .05% level of sig- nificance for 2 degrees of freedom.

Appendix E Table 2

Cumulative Frequencies of All Pupil Responses for All Teachers

	Too Fast	Too Slow	Inter-ested	Bored	Under-stand	Don't Under-stand	Good	Total
First 10 minutes	97[a] .031[b]	152 .049	757 .245	464 .151	743 .240	175 .050	675 .220	3063
Second 10 minutes	140 .039	171 .048	1073 .304	616 .174	1110 .317	281 .079	741 .209	3532
Third 10 minutes	152 .032	173 .037	1184 .253	640 .136	1229 .263	287 .061	1003 .214	4668
Fourth 10 minutes	83 .030	108 .031	820 .290	493 .178	884 .320	184 .065	788 .285	2760
Last 10 minutes	56 .031	42 .023	423 .234	264 .145	450 .249	93 .051	478 .259	1806

a. To be read as frequency of response for "too fast".
b. To be read that 3.1% of all first ten minute responses were recorded as "too fast".

28

27

Appendix E   Table 3

Comparison	r x y	$\bar{x}$	$\bar{y}$
Lecture-Interest	-.0828	.4961	.2494
Discussion-Intest	-.1591	.1406	.2494
Transition-Interest	-.2353	.3597	.2494
Lecture-Understand	+.7553[a]	.4961	.2802
Discussion-Understand	-.4977[b]	.1406	.2802
Transition-Understand	-.6516[a]	.3597	.2802
Lecture-Good	-.4934[b]	.4961	.1763
Discussion-Good	-.1178	.1406	.1763
Transition-Good	+.6433[a]	.3597	.1765

a. Significant at the 5% level, for 14 degrees of freedom.

b. Significant at the 10% level, for 14 degrees of freedom.

BIBLIOGRAPHY

1. Bird, Grace E. "Pupils Estimate of Teachers," Journal of Educational Psychology, Vol. 8, 1917.

2. Bryan, Roy C. "Pupil Ratings of Secondary School Teachers," School Review, Vol. 46, May, 1938.

3. Albert, H.P. "An Analysis of Teacher Ratings by Pupils in San Antonio Texas," Educational Administration and Supervision, Vol. 27, April, 1941.

4. Bowman, Earl C. "Pupil Ratings of Student Teachers," Educational Administration and Supervision, Vol. 20, February, 1934.

5. Gage, N.L. Handbook of Research on Teaching, Chicago: Rand McNally and Company, 1963.

6. Smith, Alden W. "Students Evaluate Teaching," U.S. Dept. HEW, OE, 1969.

7. Hedges, W.D. and MacDougall, M.A. "Recording Student Perceptions of Teacher Behavior by Means of a Student Response Monitor," The Journal of Educational Research, Vol. 58, No. 4, 1964.

8. Coats, William D. "Students Perceptions of Teachers - A Factor Analytic Study," American Educational Research Association, Washington, D.C. Paper delivered at A.E.R.A., Minneapolis, Minnesota.

9. Parekh, Jal S. "A Study of Teacher-Pupil Interaction in High School Biology Classes," Unpublished Doctoral Dissertation, Cornell University, Ithica, New York, 1965.

10. Kondo, Alan K. "The Questioning Behavior of Teachers in the Science Curriculum Improvement Study Teaching," Presented at the NARST meeting, Pasadena, California, February 7, 1969.

11. Moser, Gene W. and Feldgoise, Roberta, "Project in the Use of Interaction Analysis to Increase the Use of the Inquiry Method in the Teaching of Science," Science Project Center Report, April, 1968.

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

13. Simpson, G.G., Roe, A., and Lewontin, R.C., Quantitative Zoology, New York: Harcourt Brace and Company, 1960.