THE EFFECTS ON ACHIEVEMENT OF METHOD AND SCHEDULE OF MARKING JUNIOR SECONDARY STUDENTS' LABORATORY REPORTS

by

William Oscar Marble Jr.

B.Sc., University of British Columbia 1963

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS (EDUCATION)

in the Faculty

of

Education

SIMON FRASER UNIVERSITY

March 1977

APPROVAL

Name: William O. Marble

Degree: Master of Arts (Education)

Title of Thesis: The Effects on Achievement of Method and Schedule of Marking Junior Secondary Students' Laboratory Reports

Examining Committee

Chairman: Dr. M. Manley-Casimir

Dr. J. Martin
Senior Supervisor

Dr. P. Winne
Assistant Professor

Dr. M. McClaren
Associate Professor
Faculty of Education
Simon Fraser University
External Examiner

Date approved March 28, 1977

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis or dissertation (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this thesis for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis/Dissertation:

THE EFFECTS ON ACHIEVEMENT OF METHOD AND SCHEDULE OF MARKING JUNIOR SECONDARY STUDENTS' LABORATORY REPORTS

Author: _____

(signature)

WILLIAM O. MARBLE

(name)

Mar 28/77

(date)

ABSTRACT

In contemporary laboratory-oriented science programs, students' reports of investigations are a major component of instruction and evaluation. A question arises regarding the effects of the method of grading these reports and the frequency of the grading on students' achievement. Previous studies generally have shown partial schedules of grading to positively influence achievement relative to marking all assignments. The influence of written comments plus grades versus grades alone is far from clear. Moreover, no research has examined both factors simultaneously.

This study examined the influence of grading procedures on laboratory report marks, length of reports, and post-unit achievement over eight weeks. In a 2 x 2 factorial design, students within each of two junior-secondary biology classes were randomly assigned to receive (a) grades alone or grades plus corrective comments on (b) every laboratory report (continuous schedule) or a sample of approximately half the reports (partial schedule).

Analyses of variance showed no treatment effects on average laboratory report marks or average length. Using Gates-McInitie vocabulary subtest scores to represent general ability, analysis of covariance showed continuous schedule grading to significantly enhance achievement relative to the partial schedule ($p = .04$). Grades alone on reports produced marginal increases in achievement over grade plus written comments ($p = .12$). No interactions were statistically

significant.

This study, conducted under normal classroom conditions while maintaining rigorous experimental controls, has direct implications for teachers. Specifically, while all or almost all student assignments should be marked, extra time spent in writing corrective comments fails to yield gains in overall achievement and has no benefit to achievement in day-to-day assignments. This extra time may be more profitably used for other teaching activities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Over the past decade, junior secondary school science programs have emphasized student laboratory activities. A major product of these activities are the reports written by the students which describe experiments or projects. In these reports, data are recorded and described, observations of what occurred are listed, and explanations of observed events are offered. The report subsequently becomes one of the major sources contributing to the teacher's evaluation of students.

Marking laboratory reports generally consists of two functions: assigning a grade, and commenting on and correcting the content of the report. As the teacher's time is limited, the teacher's activities must be justified in terms of doing things in proportion to their relative contribution to promoting student learning. It is assumed that students' activities in producing laboratory reports are beneficial influences on students' achievement. Given this assumption, several critical issues bearing on teachers' use and marking of laboratory reports can be raised. Is it better to grade, comment and correct student laboratory reports or is merely grading the reports adequate? From the same theoretical perspective, one can also ask if the schedule of grading significantly influences learning?

A number of investigations have examined the effectiveness of grades and comments for enhancing student learning. The schedule of marking, whether partial or complete, has also been the subject of prior research. This study examined student achievement in a laboratory science course as a function of two features of marking: method and

schedule. More specifically, the conditions examined in a factorial design were grades with limited comments only, versus grades plus detailed comments and corrections; and grading all reports, versus grading only a portion of them.

A number of features distinguish this study from most others. Rather than using university students, this study used junior secondary students. Also, students within two classes were randomly assigned to the four treatment groups thus eliminating the problems associated with using intact classrooms. Moreover, a variety of techniques were used to maintain normal classroom conditions that were equivalent among the treatment groups. Moreover, the length of the study was nearly a full term - approximatley two months. Each of several of these items usually was absent in preceeding research studies, resulting in limitations on their findings, these limitations were circimvented in the present research.

CHAPTER II

RELATED RESEARCH

The literature review is presented in three sections. The first surveys the literature on the effects of grading and comments on achievement. The second section deals with the factor of sampling student work for grading - partial or complete. A critical survey of the research in each of these areas is provided. Finally a summary is given of the research for each factor as it applies specifically to this study. In each section an attempt was made to find a fairly significant study to provide an historical perspective on the problem. This is followed up by current research that applies to this investigation.

Grades and Comments

Most studies on grading examined the effects of grades as incentives to students to improve their performance, although the use of grades alone as incentives was rare in these studies. That is, most research examined the influence of grades combined with other factors. While most of the studies were conducted at the university level, some studies using elementary and secondary students were located.

Of the studies related to the use of teacher comments to improve performance, only a few could be found that explored the effects of written comments. All of these employed written comments in conjunction with other variables such as grades, verbal praise, and monetary reward (e.g. Houten, Hill & Parsons, 1975). Little research has been done on the efficacy of different characteristics of comments such as the type, the content, or the length of comment made.

An interesting contradiction in results regarding the effects of written teacher comments on achievement seems to indicate that the effectiveness of written comments on student performance might depend on the length and specificity of the comments. Hammer (1971) examined the effects of differential written teacher comments on student performance as measured by scores achieved on a teacher-prepared examination. The experimental treatment consisted of specific comments in addition to a grade, whereas the control group received only grades on their examination papers. Using undergraduate physics students, he found that written comments in addition to a grade had little effect in improving performance on subsequent examinations.

Hammer's study attempted a partial replication of earlier work by Page (1958) which had shown that written comments did improve performance. These earlier results seem to be coincident with the findings of others on the effects of vocal praise (Retish, 1973, Houten et al, 1975). Hammer's replication study used short written comments and the subjects were of university level. In Page's initial research the treatment employed longer comments, but was conducted with secondary school students a factor which hampers direct comparison with Hammer's study. Stewart and White (1976) after analysing the lack of success of thirteen studies which attempted to replicate Page's work, concluded that written comments were less effective at the elementary and secondary levels than at university levels. Where comments were effective, they tended to be longer in length, specific to each student, and encouraging in nature.

Although the use of grades is common in education, only recently

have their effects on students been the subject of serious study. Cullen, Cullen Jr., Hahow and Plouffe (1974), using high school students, found that when grades were used as rewards, they had a strong positive effect on the completion of assignments. The use of grades as negative incentives, i.e. marks subtracted from final grade, also resulted in a high rate of assignment completion.

Kositsky and Franken (1970) found that grades had a strong positive effect on student performance at the university undergraduate level. Their results showed that students whose assignments were graded performed at a higher level on a written assignment than students who did not receive grades. The performance of students at the graduate level has also been shown to improve when students competed for grades (Clark, 1969). Clark's work demonstrated that competition for grades on a research assignment produced significantly higher performances.

The effects of grades upon attitudes were addressed in a study using university students by Bostrom, Vlandis and Rosenbaum (1960). Students were required to write an essay defending a position contrary to their own. Grades of A and D were assigned randomly to the treatment groups. Students who received high grades showed a greater shift in attitude towards the position taken in the essay than did those who received no grade. Attitudinal effects of poor grades and no grades were indistinguishable. Another indication of the effects of grades on attitudes is the work of Bridgham (1972). He found that the opportunity for higher grades was greater in some courses than in others and that this factor influenced students' course selection. His work was done on

enrollment and grading in science courses where he found a series of significant positive correlations ($p < .05$) between the size of the particular science course enrollments and ease of grading in the course. It is unclear from the study whether the higher means for the distribution of grades necessarily implies that higher grades were easier to obtain. Perhaps selection by the student himself put brighter students in the classes or the grading system was similar but the course content was different. Another author (Dietrich, 1973) has discounted the negative effects of lower grades upon enrollment in his study of physics courses in a large number of high schools. Dietrich's results did not support the previous contention, that lower grading practices discourage enrollment. He found that physics teachers in schools with high physics enrollment tended to be more severe graders than those in schools with lower physics enrollments. While the effects of grades on attitudes do not relate directly to the factors being examined in this study, they do demonstrate the potent effect of grades on other aspects of the learning situation.

Schedules of Grading

The last group of studies relates to methods of sampling student assignments for grading. The main focus here was on the relative effectiveness of grading all assignments (a complete schedule) versus grading only some of the assignments (a partial schedule). Jenkins and Stanley (1950) extensively reviewed the effects of partial and complete schedules of reinforcement on the acquisition and resistance to extinction of learning. They found that while learning appears to be greater in the initial stages when a continuous schedule is used, over a longer period a

partial schedule is more effective in promoting retention. For the purposes of this study the preceding research provides a framework for observation that may indicate that when grades are assigned on a partial schedule, the initial learning may not be greater than on a continuous schedule, but retention might be greater after an interval of time.

Chansky (1960), found that learning occurred more rapidly when grades were given continuously than when they were given on a partial schedule. The study explored the ability of university level students to recall information when the number of correct responses acted as the dependent variable. Phillips (1965), in a study using general psychology students, examined the effects of partial grading on 'essay-type' achievement tests. While his experiment lacked controls and random assignment, he found that a partial grading schedule tended to increase the motivation for more thorough study. Thus, this method of grading seems to provide a means for a teacher to reduce marking load while maintaining learning, or an atmosphere which promotes learning.

Research that bears more directly on the present study is provided by Cohen (1971). He studied the effects of evaluating student laboratory reports on different grading schedules in grade nine. The results showed that the number of acceptable laboratory reports submitted was unrelated to the grading schedule, i.e. continuous or partial. Since the students in the study were not randomly assigned to the treatments and were of above average intelligence, a generalization about the time-saving value of partial grading is tenuous. Further, since Cohen collected one laboratory report per week from each student, the question might be raised

as to whether the lack of significant difference in achievement at the end of the study between the partial and continuous groups was due to the rather long interval of over one week between submitting the report and receiving the grade. Perhaps a shorter interval would have improved the effect of the grades. The study does, however, raise some important questions about the practice of grading each student's report in science classes.

Other research (Rosenfeld, 1972) indicated that students' level of intelligence is closely related to the effectiveness of regular classroom reinforcement and also to the degree of improvement noted. High I.Q. students improved their performance on arithmetic tests significantly more ($.01 \leq p \leq .05$) than other I.Q. groups when classroom reinforcement included grades and teacher praise. Specifically, more intelligent students showed the greatest increases in performance, under most types of incentives.

Summary

In summary the evidence suggests that grades have a consistent, powerful and positive effect on student achievement. There is some suggestion, however, that the ability of the student and the type of grade received, i.e. mark versus mark plus comments, for an assignment or test bears on the effectiveness of grades in improving achievement or performance. The precise effect of the length and specificity of written teacher comments on student achievement has not been properly clarified. Studies on schedules of grading, i.e. partial or complete, suggest that

properly applied partial grading schedules can benefit student achievement as much as continuous grading practices. In this study, we have attempted to clarify the effects of the two variables discussed in this review, namely, marking schedules and written comments. The written comments have a high specificity to overcome the limitations found in other studies. The data gathered in this study on the effects of types of scheduling will hopefully resolve some of the conflicting results in this area.

CHAPTER III

METHOD

Subjects

The target population for this study consisted of students enrolled in laboratory-oriented junior secondary science programs where the marking of written reports by the teacher served the functions of reinforcing student behaviour and providing information for evaluation.

The sample for the study consisted of a total of 53 heterogeneously grouped grade eight science students each of whom participated in one of two separate science classes. The students came from a wide range of socio-economic backgrounds, with a fairly high percentage (approximately 40%) coming from single-parent and recent immigrant homes. The school operated on a rotating eight block cycle with five periods each day. In this way, students received five 1-hour science classes over every eight school days. The two classes were scheduled in consecutive blocks. Students had been assigned to their classes by a computer prepared timetable. Prior to this study they had spent approximately five months in the science course within which this study was conducted. In the laboratory, the students worked in pairs at 1 of 6 stations. Each station contained the working area for two pairs of students. Whenever a class size exceeded 24 members, provision was made for some students to work as a group of three.

The existing pairs of students in both classes were randomly assigned to 1 of the 4 treatment groups in the study, so that all four treatment groups were represented in each class. Thirteen students participated in 3 of the 4 groups. The fourth group contained 14

students.

Background

The general format for each class period involved three aspects. First, it began with a short discussion of the previous laboratory experiment. This was followed by preparation for the next laboratory exercise, including any special instruction about the procedure and the write-up of the report. There were usually about 35-40 minutes remaining in the period for the students to complete the experiment and the report associated with it. An example of the general outline for laboratory reports is presented in Table I.

TABLE I

General Outline for Laboratory Reports

Student's Name
Partner's Name
Problem: Copy the title from the text.
Procedure: Answer the questions in heavy print in the procedure.
Include data tables and graphs if required.
Use full sentences.
Conclusions: Answer the assigned questions at at the end of the exercise.

Prior to the study itself, laboratory reports were collected from each student on the average of one per week. All students handed in a report on the same laboratory when requested. The students were not aware that a report would be collected until the last fifteen minutes of

the laboratory period, at which time the teacher announced that the reports were to be handed in. The reports were graded on a scale of zero to five with zero assigned only to those students whose reports were not handed in. The reports also received comments indicating whether students recorded particularly good observations or reached astute conclusions. Careless work habits, such as incomplete work or illegible writing, were also noted. Errors in arithmetic, spelling and sentence structure were corrected. Encouraging comments of a more general nature such as "A very well written lab" and "This is a good example of careful work" were used frequently when appropriate.

Materials

The students spent a total of 20 hours in the laboratory during the study. Instruction for all students was delivered by the same teacher, who was also the researcher. The laboratory exercises were taken from the text: Introducing Science Concepts in the Laboratory (M. Schmid, editor, 1973), specifically Unit Two: "Living Things Detect and Respond to Stimuli" (pp. 51-100). While some of the experiments outlined in the text were followed exactly, others were modified considerably by the experimenter-teacher. The unit consists of investigations of plant and animal responses to stimuli. Plants were examined for their responses to light and gravity while earthworms and a selection of arthropods were studied for their responses to light, moisture, and other stimuli. The unit consisted, in part, of experiments where various types of seeds were germinated and allowed to develop into seedlings.

Treatments and Procedure

Students were randomly assigned within each class to one of the four treatment groups. All groups were required to hand in all laboratory reports in order to control the possible reinforcing effect that handing in reports might have on students. The "complete sampling but grades only" group (C,G) received only a grade for every report handed in. The "complete sampling with grades and comments" group (C,G+C) received written comments and corrections on all reports in addition to grades. The "partial sampling but grades only" group (P,G) was given grades only for a sampling of the reports handed in whereas the "partial sampling with grades and comments" group (P,G+C) received the written comments along with grades. For this study, partial sampling consisted of marking every other one of the 12 reports. With only minor exceptions, these were alternate reports. The following factorial matrix (Figure 1) depicts the four treatment groups:

FIGURE 1

		Method of Marking	
		Grades only	Grades plus Comments
Method of Sampling	Complete	C,G	C,G+C
	Partial	P,G	P,G+C

As some laboratory exercises required several periods to complete, students were often recording observations and data on several experiments during a single period. When an experiment was finished the entire class

handed in their reports.

All reports, marked and unmarked, were returned to the students at the beginning of the following science period. While all reports were graded for the purposes of analysis, students in the groups on the partial schedule were informed of their grades on only half of the reports.

Grading Procedure

Standardization of the grades given to students was obtained by specifying the exact content required in each laboratory report to receive a grade of five, four, three, two or one. Zero was reserved only for those students who failed to hand in a report. Before comments were made on students' reports in the G+C groups, a list of student errors was constructed based upon the past experience of the teacher. Each error was then matched with a single written comment to ensure that the same comment would be used for the same error on all reports with that error which received comments. Errors in spelling and arithmetic were corrected on all reports from all groups. The only additional notation used on all reports of the C,G and P,G groups was a "?" for missing material. An example of how the grading and comments were standardized is given in Table II for the first laboratory in the unit. The comments applied only to the C,G+C and the P,G+C groups. These comments were specific to the first laboratory and are included as an example of the type of comment made. Each laboratory exercise had a separate set of comments, though there were comments common to a number of laboratories.

TABLE II

Errors, Standardized Comments, and Grading Scheme for Exercise 2-1: "Planting an Experimental Garden" (Schmid, p. 55-58)

Errors	Comments/Corrections
1. Problem not given	"See title in the text."
2. Incomplete sentences	"Use full sentences please!"
3. Sketches required in step 9: (a) wrong size	"Is this really 4X the actual size?"
(b) unlabelled	List the missing parts with a question mark beside the list.
(c) mislabelled	Use arrows to indicate the correct part.
(d) without sketch magnification	"Sketch magnification?"
4. Incorrect prediction of mature names of plumule, radicle and hypocotyl	"Which part looks like (leaves)? (stems)? (roots)?

Grading Scheme

5 - All observations included and correct. Drawings correct size with all labels correctly placed.

4 - One or two minor errors in observations and/or drawings (such as 2, 3(c)(e), and 4 above).

3 - One major error such as 3(a), (d) above) or up to three minor errors in observations and/or drawings.

2 - More than one major error or three minor errors in observations and/or drawings.

1 - Report incomplete.

0 - Report not handed in.

A copy of a marked report complete with comments is included in Appendix D.

Test Instruments

Before the study began all students completed the vocabulary section of the Gates McInitie Reading Test, Survey E, Form 2. This measure was used as an indicator of general mental ability. The test has a split-half reliability coefficient corrected by the Spearman-Brown formula of +.89, a test-retest reliability (Pearson coefficient) of +.80 and a correlation with the Lorge-Thorndike Verbal I.Q. of +.74 for the grade eight students participating in this study.

The main dependent variable, student achievement, was measured by a teacher-prepared test containing 52 multiple choice and matching items. The test is reproduced in Appendix A. The selection of the test items and their revision had been an ongoing process over a four year period. The test requires approximately 45 minutes to complete and can be easily administered within a regular class period. Measures of central tendency and variation, calculated from the scores of 47 grade eight students who completed the test in a previous year appear in Table 1 of Appendix B. The internal consistency coefficient of the test for this earlier administration was +.72 using the Kuder-Richardson Formula 21. The standard error of measurement was 2.9. A logical analysis of the percentage of items at each level of Bloom's (1970) cognitive taxonomy is given in Table III.

TABLE III

Taxonomic Levels of Achievement Test Items

Level	Percentage of Items at Level
Knowledge	42
Comprehension	12
Application	25
Analysis	4
Evaluation	17

Test-retest reliability (Pearson coefficient) calculated from the scores of six grade eight students at two sittings 48 hours apart was +.92. The scores are given in Table 2 of Appendix B. The split-halves reliability coefficient, corrected by the Spearman-Brown prophecy formula and calculated from the scores from another sample of 16 grade eight students was +.71. The scores are given in Table 3 of Appendix B. Concurrent validity was estimated using the test scores and an independent ranking of the same students by their regular teacher. The Spearman rank-order correlation was +.61 ($p < .001$). The scores and ranks are found in Table 4 of Appendix B.

The marking time for each report was recorded using a stopwatch. This recording included the time to assign the grade and, for the group also receiving comments, to grade and write comments. The length of each report to the nearest tenth of a page was also recorded.

When the reports were returned to the students at the beginning of

the next period, approximately 10-15 minutes were spent in a general discussion of the experiment. Specific errors were not discussed to avoid contamination of the comments versus the no comments treatment. The teacher encouraged the students to offer their observations and ascertain whether some general conclusions about the experiment could be made. This discussion was often a "lead-off" into the next piece of laboratory work. A record was kept of any special events or content in the discussion that occurred. A record of student absences and missing laboratory reports were kept prior to and throughout the study.

In order to provide data on characteristics of the marking scheme in use before the study began, the following procedure was used. An unmarked report was collected from the students for a laboratory done previous to the study. A random sampling of 8 reports was taken and photocopied. The originals were then marked. Approximately two weeks later the unmarked copy was marked and the grade, types of comments, marking time, and length of report were compared between the first and second markings. This same procedure was adopted during the experiment itself. Approximately mid-way through the unit, another sampling of 8 reports for a single laboratory was copied before marking and then following marking. These copies were put away until the end of the study when the previously unmarked copies were marked. These were compared to the report which had been marked earlier to check marking consistency during the study.

At the end of the unit, the achievement test was given as a posttest to all students in the study. They were given five days

notice for the test and told that the results would be used as part of the evaluation on the upcoming report cards.

A breakdown by laboratory assignment was made of the questions on the achievement test. This was to check whether the outcomes of particular laboratories or the reports marked only for the continuous treatment groups received greater emphasis on the test. The results given in Appendix C show that 4 laboratories out of the 8 accounted for 55% of the questions and the laboratories marked only for the continuous group accounted for 45% of the questions. The balance of the questions, then, was not reasonable.

The achievement test was administered to all students in the two classes immediately prior to this study as an indicator of extant knowledge about the curriculum. The test and the results were not discussed with the students except to indicate that the results would be helpful to the teacher in judging how much students knew before the study began. The test was not returned.

The second dependent variable was the grade received by the students on their laboratory reports. The procedure for grading these reports is described in a later section.

CHAPTER IV

RESULTS

Missing data in this study were the result of student absence or reports not handed in. As noted earlier, a report not handed in received a grade of zero. If a student was absent no grade was assigned. Obvioulsy for both of these situations marking time or report length data were not recorded.

As mentioned previously, in order to check the general consistency of marking, a random sample of 8 reports was selected, marked and remarked two weeks later without reference to the first marked report. This procedure was repeated before and during the study. The results are presented in Table IV. The Pearson correlation coefficient between the grades for the two time blocks was .92 on the reports collected preceeding the study and .89 on the reports collected during the study. Similarly the correlation coefficients for the marking times were .59 before and .63 during the study. These results indicate that the grades received were very consistent. This reliability of the teacher's marking consistency over time, combined with the grading scheme of the type given in Table II helped to minimize possible extraneous variation in the laboratory grade scores.

TABLE IV

Consistency of Grading and Marking Time

Preceding and During Study

	Preceding Study				During Study			
	First Marking		Second Marking		First Marking		Second Marking	
	Grade	Mark Time	Grade	Mark Time	Grade	Mark Time	Grade	Mark Time
	3	48	3	45	5	29	5	42
	4	95	4	46	4	40	3	45
	4	57	4	41	5	90	5	50
	3	115	4	63	5	15	4	35
	4	105	4	88	3	20	3	45
	2	87	2	85	3	90	3	67
	2	84	3	56	5	68	5	51
	2	72	2	58	5	69	5	87
Mean	3.0	82.9	3.0	60.3	4.4	52.6	4.1	52.8

Note: Maximum grade = 5. Mark time in seconds

To examine the possibility that the treatment groups differed in general ability or on prior knowledge of the curriculum taught during the study, separate one-way analyses of variance were performed on the Gates-McInitie Reading Test scores and the pretest achievement scores. The results of these analyses are given in Tables V and VI. In addition the means and standard deviation for the scores on the Gates-McInitie Reading Test and the Pretest for each treatment cell are given in Tables VII and VIII. No statistically significant differences were found for either variable. The four treatment groups were judged to be equivalent for these measures, as would be expected due to the semi-random assignment procedures by which pairs of students were randomly assigned

to treatments within each class. Each pair of students received the same treatment which was designed to reduce the chance of bias due to communication between partners during the laboratory activities.

TABLE V

One-way Analysis of Variance

of Gates-McInitie Reading Test Results

by Treatment Groups

Source	SS	df	MS	F	p
Between Groups	134.28	3	44.76	1.34	.10
Within Groups	1642.04	49	33.51		
Total	1776.32	52			

TABLE VI

One-way Analysis of Variance

of Pretest Results

by Treatment Groups

Source	SS	df	MS	F	p
Between Groups	99.65	3	33.22	1.39	.10
Within Groups	1173.68	49	23.95		
Total	1273.33	52			

TABLE VII

Breakdown by Treatment Groups for

Gates-McInitie Reading Test

		Method of Marking		
		G	G+C	Total
Schedule of Marking	C	18.85 6.76	23.08 6.30	20.88 6.77
	P	21.15 4.63	22.15 5.11	21.65 4.81
Total		19.96 5.84	22.61 5.64	21.26 5.84

Note: Means are upper numbers, standard deviations are low numbers. Maximum score is 32.

N = 13 for all groups except C,G (N = 14).

TABLE VIII

Breakdown by Treatment Groups

on the Pretest

Schedule of Marking	Method of Marking: G	Method of Marking: G+C	Total
C	14.36 4.55	17.23 5.06	15.74 4.93
P	14.38 3.82	13.61 5.92	14.00 4.90
Total	14.37 4.12	15.42 5.70	14.89 4.95

Note: Means are upper numbers. Standard deviations are lower numbers. Maximum score is 52. Highest score obtained is 29. N = 13 for all groups except C,G (N = 14).

The mean pretest and posttest scores for each treatment group are compared in Figure 2. The graph shows that all treatment groups markedly improved their knowledge of the curriculum taught during the study. Accompanying the means are the standard errors of measurement for both the pretest and the posttest scores. Table IX shows the t-test results and level of significance contrasting the pre-and post-test scores for each treatment group. These verify that learning did occur during the study.

FIGURE 2

A Comparison of Mean Pretest and Posttest Scores by Treatment Groups

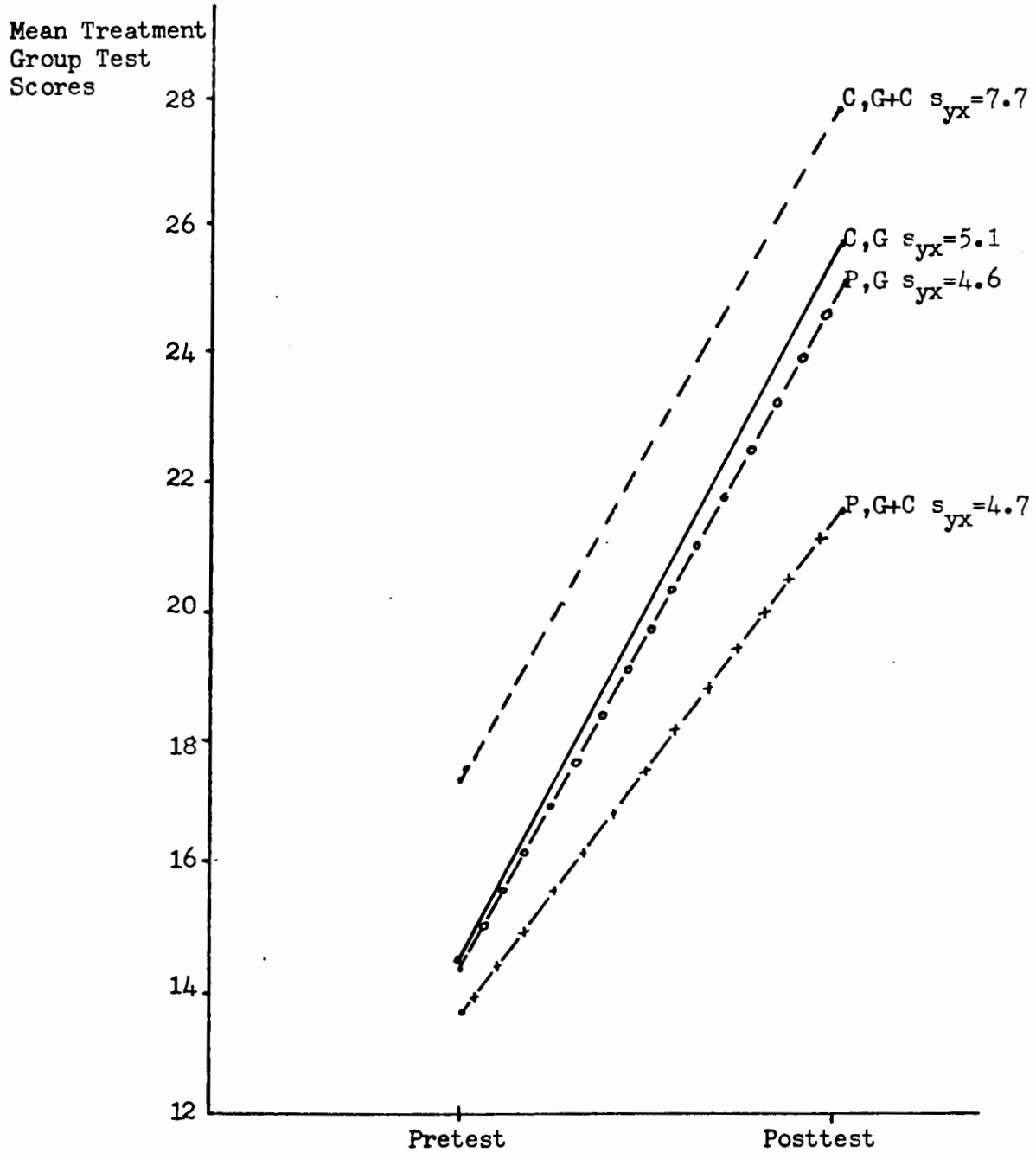


TABLE IX

Summary of t-test Results Between the

Pretest and Posttest Means

Schedule of Marking	Method of Marking: G	Method of Marking: G+C
C	$t = 3.92$ $p < .001$	$t = 3.74$ $p < .001$
P	$t = 5.53$ $p < .001$	$t = 7.84$ $p < .001$

As stated previously the purpose of this study was to examine the effects of type and schedule of marking laboratory reports on student achievement on laboratory reports as well as end-of-unit achievement. Figure 3 shows the mean laboratory grade received by each treatment group for each laboratory. It is important to note again that students in the P groups had all reports marked but grades were returned to them on about one-half of the reports (i.e. laboratories 3, 4, 7, 8, 11, 13). Little difference between the groups is indicated, with the possible exception of the P, G+C group which had the lowest mean grade on about 60% of the laboratories. The mean laboratory grades for each treatment group are given in Table X. A two-way analysis of variance (Table XI) shows that the difference among group laboratory mean scores was not statistically significant.

FIGURE 3

A Comparison of Mean Report Grades by Treatment Groups for Each Laboratory

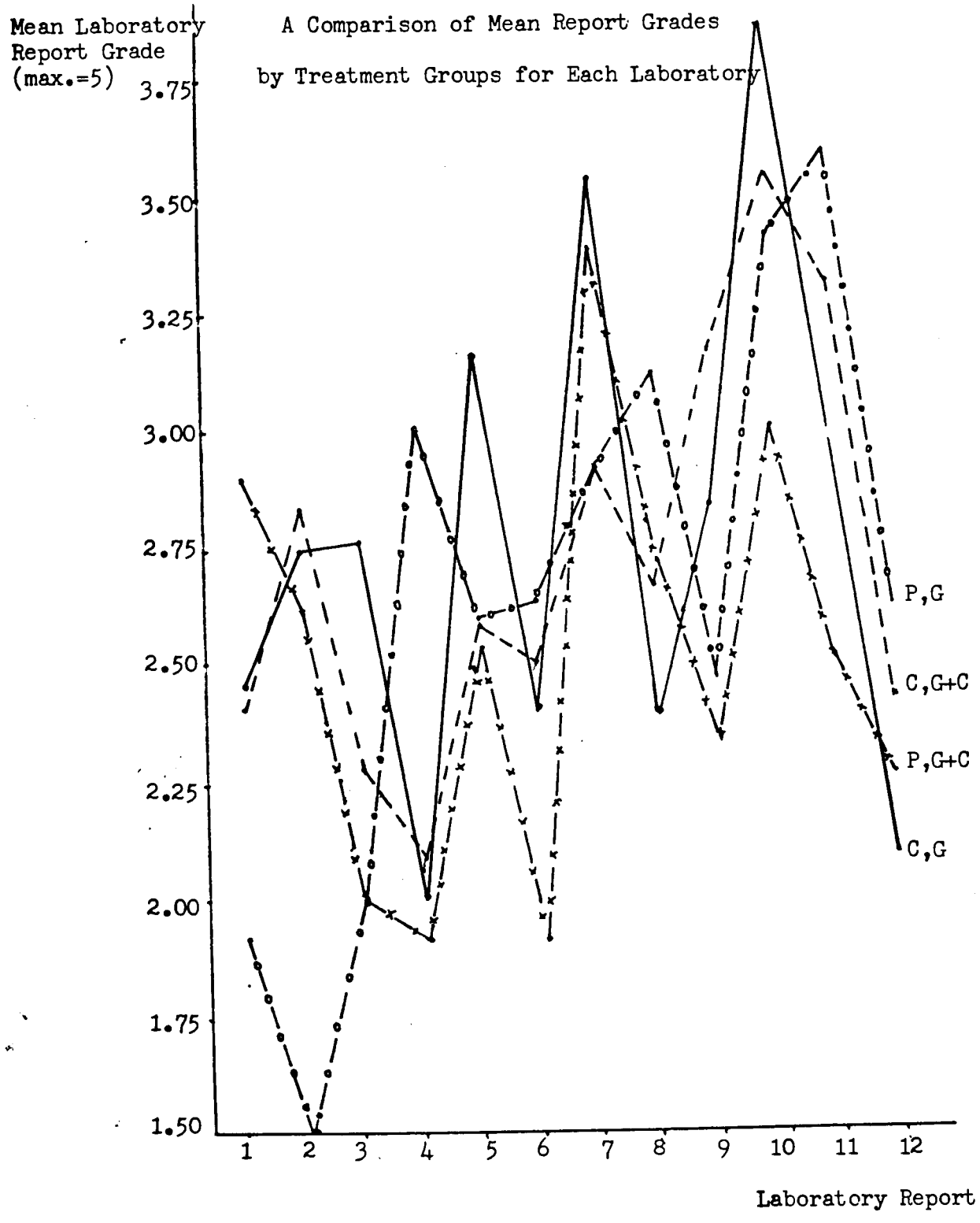


TABLE X

Breakdown by Treatment Group

of Laboratory Grades

Schedule of Marking	Method of Marking: G	G+C	Total
C	2.7 1.2	2.6 1.4	2.7 1.3
P	2.8 1.0	2.5 1.0	2.6 1.0
Total	2.8 1.1	2.6 1.2	2.7 1.2

Note: Means are upper numbers, standard deviations are lower numbers. Maximum score is 5.

TABLE XI

Two-way Analysis of Variance

of Average Laboratory Report Grade

by Treatment Groups

Sources	SS	df	MS	F	p
Main Effects:	.23	2	.11	.10	.99
Schedule of Marking	.14	1	.14	.13	.99
Method of Marking	.09	1	.09	.08	.99
2-Way Interactions	.02	1	.02	.02	.99
Explained	.23	3	.08	.07	.99
Residual	54.20	49	1.11		
Total	54.43	52	1.05		

The comparison of the mean report lengths in Figure 4 shows wide variation between laboratories but little variation between the treatment groups within each report. The length of a student's report does not seem to have been affected, however, in any significant way according to the type of marking schedule (see Table XII). A two-way analysis of variance (see Table XIII) on average rrport length for the four treatment groups confirmed this observation.

FIGURE 4

A Comparison of Mean Report Length
by Treatment Groups for Each Laboratory

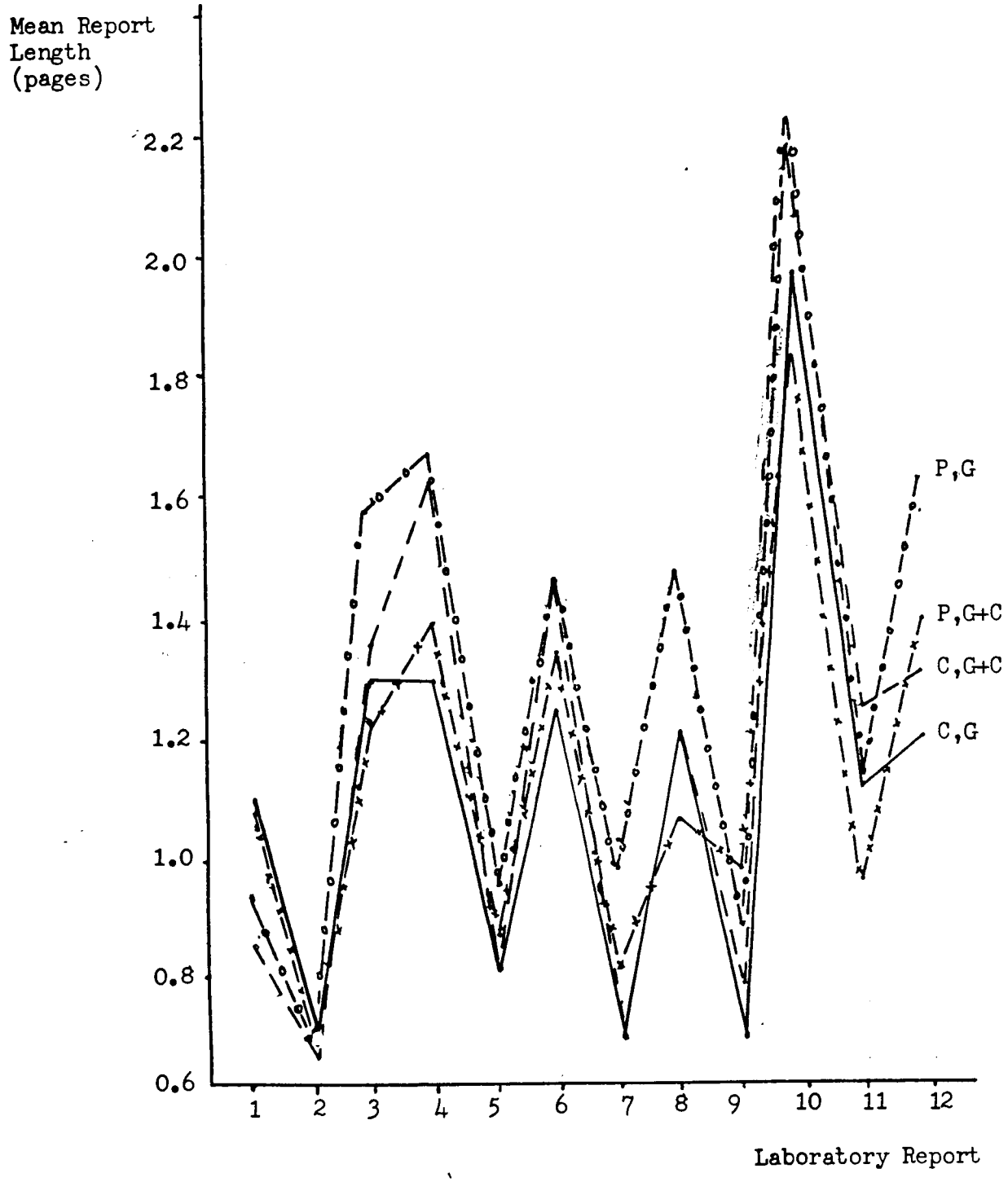


TABLE XII

Breakdown by Treatment Group

of Average Report Length

Schedule of Marking	Method of Marking: G	G+C	Total
C	1.1 .30	1.2 .38	1.2 .34
P	1.3 .37	1.1 .32	1.2 .34
Total	1.2 .33	1.2 .34	1.2 .34

Note: Means are upper numbers, standard deviations are lower numbers.

TABLE XIII

Two-way Analysis of Variance

of Average Laboratory Report Length

by Treatment Groups

Sources	SS	df	MS	F	p
Main Effects:	6.99	2	3.49	1.84	.17
Schedule of Marking	2.70	1	2.70	1.42	.24
Method of Marking	4.41	1	4.41	2.32	.13
2-Way Interactions	1.12	1	1.12	.59	.99
Explained	8.11	3	2.70	1.42	.25
Residual	93.20	49	1.90		
Total	101.31	52	1.95		

Two of the measures kept during the study were pupil attendance and reports not handed in. Table XIV summarizes the results by treatment group. An examination of the table would indicate that attendance did not vary significantly across the groups. In the C,G and C,G+C treatment groups, the large number of reports not handed in was the result of two students in the C,G group and three in the C,G+C group who failed to hand in more than two reports. These seem to be individual differences among students rather than effects attributable to the treatment conditions per se.

TABLE XIV

A Summary of Student Attendance

and Reports not Handed In

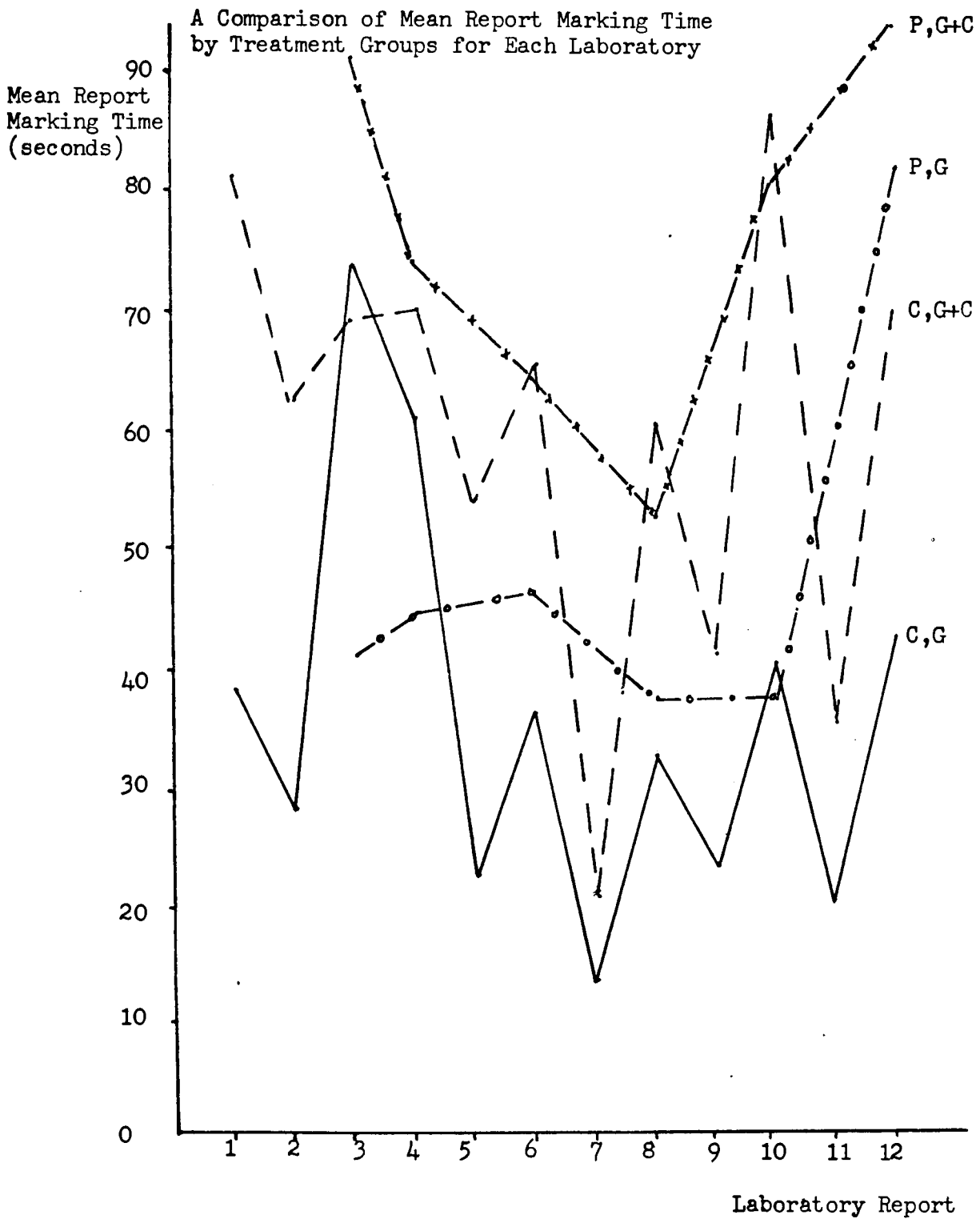
by Treatment Groups

Treatment Group	Students Present	Student Absence	Reports Not Handed In
C, G	153	15	10
	12.8	1.3	.8
C, G+C	137	19	18
	11.4	1.6	1.5
P, G	144	12	4
	12.0	1.0	.3
P, G+C	137	19	6
	1.4	1.6	.5

Note: Lower numbers are means. "Students Present" is the product of the number of students in the treatment group and the number of periods.

Figure 5 is a breakdown by laboratory of the mean marking time for the four groups. Table XV is a breakdown by treatment group of average marking time over the unit. The marking time consisted of three factors: (1) reading the report, (2) placing a grade on the paper and for the G+C groups, (3) placing written comments on the paper. As the P groups had only certain reports marked it was necessary to extrapolate their lines

FIGURE 5


A Comparison of Mean Report Marking Time by Treatment Groups for Each Laboratory
Mean Report Marking Time (seconds)
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
Laboratory Report
P,G+C
P,G
C,G+C
C,G

through the laboratories where they actually received grades and/or comments on their reports. It is interesting to note that the continuous groups received somewhat more teacher marking time than did the partial group.

TABLE XV

Breakdown by Treatment Group

of Average Marking Time in Seconds

		Method of Marking		
		G	G+C	Total
Schedule of Marking	C	36.21	59.98	48.10
	P	35.63	49.44	42.54
	Total	35.92	54.71	45.32

For the analysis of covariance used to test for differences among group means on the achievement posttest, the statistical assumptions for using the Gates-McInitie Reading Test scores alone, the pretest alone, or both, as covariates for adjusting the posttest means were examined. The procedure used was to statistically test the increase in the variance of the dependent variable accounted for by forcing different sequences of entry of the predictor variables in a multiple regression equation predicting the achievement posttest scores. The examination of those results (see Table XVI) suggested that the most accurate analysis would be achieved by using the Gates-McInitie Reading

scores only as covariate. This was because using the pretest alone was unsatisfactory relative to using the Gates-McInitie alone and when the pretest was used in conjunction with the Gates-McInitie, the adjustment was not statistically improved as judged by the F-statistic testing the standardized partial regression coefficient for the pretest following adjustment by the Gates-McInitie. The examination indicated, also, that the interaction effects between the covariate and the treatment effects were not statistically significant and therefore the the assumptions of homogeneity of regression had been satisfied. In the analysis of covariance (Table XVII) the difference attributable to the main effects demonstrated a significant difference for the method of marking. The adjusted multiple classification analysis of means demonstrated the continuous schedule (C) to be superior to the partial schedule (P). While differences in the method of marking failed to reach traditional levels of statistical significance, grades only (G) appeared to be better than grades plus comments (G+C). Table XVIII shows the group means adjusted for the covariate. Interaction effects of the two treatment conditiond were not observed.

TABLE XVI

Percentage of Posttest Variation

Predicted by Independent Variables

Combination	Predicted Posttest Variation
Pretest + Treatment Groups	28.6%
Gates-McInitie + Treatment Groups	41.9%
Gates-McInitie + Pretest + Treatment Groups	43.3%

TABLE XVII

Analysis of Covariance of the Posttest Results

by Schedule and Method of Marking

With the Gates-McInitie Reading Scores as Covariates

Sources	SS	df	MS	F	p
Covariates: Gates-McInitie	828.03	1	828.03	24.96	.001
Main Effects:	235.29	2	117.65	3.55	.04
Schedule of Marking	147.18	1	147.18	4.44	.04
Method of Marking	83.71	1	83.71	2.52	.12
2-Way Interaction: Schedule and Method	38.91	1	38.91	1.17	.29
Explained	1102.23	4	275.56	8.31	.001
Residual	1526.27	46	33.18		
Total	2628.50	50	52.57		

TABLE XVIII

Posttest Group Means Adjusted for Covariate

Schedule of Grading		Method of Marking	
C	26.63	G	26.19
P	23.23	G+C	23.56

To briefly review, the results of this study were that: (1) the variation in average report length was a result of the pattern of individual laboratories and exposure to the unit as a whole rather than any one of the treatment groups, (2) variation in average report grades seems to be due in part to the treatment received, even after students' general ability had beencontrolled statistically. The question as to how great an influence the treatments had on a student's learning over the unit are found in the analysis of covariance. The results showed that a great deal of the difference in posttest scores are attributable to reading ability and general knowledge. However, statistically significant differences were observed as a function of whether the student received grades on a partial or continuous schedule. The differences for the method of grading were not as statistically reliable. There were no significant interaction effects between the treatment groups.

CHAPTER V

DISCUSSION

The results of this study indicate that students in a laboratory-oriented program learned more in an eight-week instructional unit when the schedule of marking their laboratory reports was continuous rather than intermittent. These findings, while supported by some research (e.g. Chansky, 1960), contradicted a number of previous studies which found that a partial schedule of marking had a greater positive influence on student achievement (e.g. Phillips, 1965; Cohen, 1971). Table XIX compares the important characteristics of the present study with those of the parallel studies by Phillips, Cohen and Chansky.

TABLE XIX

Characteristics of Research

Examining Scheduling

	Treatment and Scheduling Frequency	Dependent Variable	Treatment Length	Subject Characteristics
Phillips (1965)	P - 50% C - 100%	Objective Midterm & Final Exam	12 weeks	University Students
Cohen (1971)	P - 25% C - 100%	Quality of Reports	18 weeks	Above avg. ability grade 9 students
Chansky (1960)	P - 50% C - 100%	Objective final test	3 weeks	University students
Marble (1977)	P - 50% C - 100%	Objective final test	8 weeks	Average grade 8 students

While the number of pertinent studies on marking schedules are few, treatment length appears to have an influence upon the effectiveness of marking schedules in improving achievement. Specifically, the research studies which found a continuous schedule to have a more positive effect had a somewhat shorter treatment period (3 - 8 weeks), whereas a partial schedule proved to be more effective when the treatment extended over a longer period (12 - 18 weeks). While the Chansky study (1960) concurred with the results of this study, little other evidence is availible to indicate what may have caused the differences in achievement compared to the Phillips and Cohen studies. Future studies which examine different levels of scheduling frequency and treatment length are clearly needed. Would the greater relative improvement shown by the continuous group in the present study continue over a longer treatment period, or are such differences between partial and continuous marking schedules a direct function of brevity of treatment?

A final consideration is the ability level of the student subjects. In this study the mental ability of the student subjects was kept statistically constant in each treatment cell. Perhaps performance under a continuous or a partial schedule might be a function of the abilities of the student, although the acceptance of the assumption of homogeneity of regression slopes in the analysis of covariance indicated that this was not the case in the present study. Future research may shed light on the possibility of this and other aptitude-treatment interactions.

For the second main independent variable in this study, grades only resulted in greater improvement in learning than grades plus written

comments. However this difference was not as statistically reliable as the one attributable to variation in the marking schedule. The literature reviewed is again, inconclusive on this question. Some studies concur with htis last one that written comments do not significantly improve student achievement (Hammer, 1971). Other studies, however. found written commentsto be somewhat more effective than grades alone (Stewart & White, 1976). Table XX provides a comparison of the characteristics of studies of method of marking.

TABLE XX

Characteristics of Research

Examining Method of Marking

	Treatment Groups	Dependent Variable	Treatment Length	Subject Characteristics
Hammer (1971)	G G+C	Final examination	very short	University students
Stewart & White (1976)	G G+C C	Objective final achievement test	6 weeks	Mixed ability grades 5 - 7
Marble (1977)	G G+C	Objective final test	8 weeks	Mixed ability grade 8

The studies reviewed earlier (Hammer, 1971; Stewart & White, 1976) and the present one used basically two treatments; (a) grades alone and (b) grades plus some form of written comments. In all cases the dependent variable was a final achievement test, (generally objective). The length

of the treatments, however, varied considerably frim very brief (Hammer, 1971) to 6 - 8 weeks (Stewart & White, 1976). Again the ability range and age level of subjects was wide; ranging from university level in Hammer's study (1971), normal grades 5 - 7 in Stewart and White's (1976) work, to randomly assigned grade 8 students in the present study.

Some evidence has been presented (Stewart & White, 1976; Hammer, 1971) which seems to indicate that where comments were more effective they tended to be longer, more specific, generally encouraging, and used with students in post secondary institutions. In this study an attempt was made to provide specific, corrective types of comments as well as comments of a constructive encouraging nature, but the previously mentioned features of specificity and length of comment were held constant for those students in the grades-plus-comments treatment group. Experiments which manipulate the attributes of length and specificity in a factorial design with several levels of such a factor would provide more complete information as to the effects of the variables on student achievement.

Nevertheless, the effects of written comments on the quality of day-to-day assignments was rather disappointing. One might question whether the effectiveness of a written comment depends upon the grade level at which it is used (i.e. elementary, secondary or post-secondary), or the ability-level of particular students. Though the question was not addressed by this study, perhaps higher-order comments might prove more effective than those of a lower-order. As mentioned elsewhere (Rosenfeld, 1972), the effectiveness of regular classroom reinforcement is closely related to the students' level of intellectual ability. Perhaps method of marking is similarly related. One might also examine

whether student perception of the comment plays a role in determining its effect. Perhaps students with lower self-concept respond better to encouraging comments than do those with higher perceptions of themselves.

Having considered the effects of method of marking and sampling of reports to be marked on student learning outcomes as measured by a teacher made test, this study has also shown that neither treatment, method of marking or sampling, significantly improved the grades received on laboratory reports (Table XI) or increased the lengths of the reports (Table XIII). One of the main reasons a teacher uses written comments is to provide relatively immediate feedback to a student about the strengths and weaknesses of his work and to encourage him to strive to improve the quality of that work in the future. The ineffectiveness of the grades-plus-comments treatment in this study suggests that other means should be explored to fulfill this objective. It might for example, be interesting to manipulate the importance of the laboratory grade component in the overall evaluation of the student's achievement.

By way of a summary, in the light of these data a teacher wishing to maximize student learning and minimize the amount of time spent marking student assignments would seem to profit more by collecting every report and grading it without written comments, providing that the instructional period was relatively short. If a teacher did decide to use written comments, positive effects might be colsely related to the care taken to use comments specific to a particular content situation rather than using comments which are general in nature (Stewart & White, 1976). The effectiveness of this treatment might be greater if used with students of higher ability over a relatively short treatment period.

BIBLIOGRAPHY

Bloom, Benjamin S., Hastings, J. Thomas and Madaus, George F. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill, 1971.

Bostrom, Robert N., Vlandis, John W., and Rosenbaum, Milton E. "Grades as Reinforcing Contingencies and Attitude Change." Journal of Educational Psychology, 52, pp. 112-115, 1961.

Bridgham, Robert "Ease of Grading and Enrollment in Secondary School Science, II A Test if the Model." Journal of Research in Science Teaching, 9, no. 4, pp. 331-343, 1972.

Chansky, Norman M. "Learning: A Function of Schedule and Type of Feedback." Psychological Reports, 7, p. 362, 1960.

Clark, D. Cecil. "Competition for Grades and Graduate-Student Performance." Journal of Educational Research, 62, no. 8, pp. 351-354, 1969.

Cohen, Ronald D. "Evaluationof Student Laboratory Reports Under a Schedule of Partial Reinforcement." Journal of Research in Science Teaching, 8, no. 2, pp. 185-189, 1971.

Cullen, Francis T., Cullen, John B., Hayhow, Van L., and Plouffe, John T. "The Effects of the Use of Grades as an Incentive." Journal of Educational Research, 68, pp. 277-279, 1974.

Dietrich, Don. "Grading Practices of High School Physics Teachers: A Contributing Factor to Declining Enrollment in Physics?" Science Education , 57, no. 1, pp. 25-29, 1973.

Hammer, Bernard. "Grade Expectations, Differential Teacher Comments, and Student Performance." Journal of Educational Psychology, 63, no. 5, pp.454-458, 1972.

Houten, Ronald Van Hill, Sharon, and Parsons, Madeline. "An Analysis of a Performance Feedback System: The Effects of Timing and Feedback, Public Posting and Praise Upon Acedmic Performance and Peer Interaction." Journal of Applied Behavioural Analysis , 8, no. 4, pp. 449-507, 1975.

Jenkins, W. O. and Stanley, J. C. Jr. "Partial Reinforcement: A Review and Critique." Psychological Bulletin, 47, pp. 193-234, 1950.

Kositsky, Nate and Franken, R. E. "Effects of Grading and Rewrite Procedures on Acedemic Performance." Psychological Reports, 27, pp. 244-246, 1970.

Page, E.B. "Teacher Comments and Student Performance: A Seventy-four Classroom Experiment in School Motivation." Journal of Educational Psychology, 49, pp. 173-181, 1958.

Phillips J. L. Jr. "An Application of Intermittent Grading." The Clearing House, 39, pp. 305-306, 1965.

Retish, Paul M. "Changing the Status of Poorly Esteemed Students Through Teacher Reinforcement." Journal of Applied Behavioural Science, 9, pp. 44-50, 1973.

Rosenfeld, George Walker. "Some Effects of Reinforcement on Achievement and Behaviour in a Regular Classroom." Journal of Educational Psychology, 63, no. 3, pp. 189-193, 1972.

Schmid, M. ed. Introducing Science Concepts in the Laboratory, California: Prentice-Hall, 1973.

Stewart, Linda Graves, and White, Mary Alice. "Teacher Comments Letter Grades, and Student Performance: What Do We Really Know?" Journal of Educational Psychology, 66, pp. 488-499, 1976.

APPENDIX A

ACHIEVEMENT TEST USED AS PRE AND POSTTEST

APPENDIX A

Achievement Test Used as Pre and Posttest

SCIENCE 8 BIOLOGY

1. A phototropism is a response to: (1) food (2) gravity (3) light (4) touch (5) water

2. A thigmotropism is a response to: (1) food (2) gravity (3) light (4) touch (5) water

3. A geotropism is a response to: (1) food (2) gravity (3) light (4) touch (5) water

4. The presence of light affects seedlings by: (1) increasing rate of growth. (2) decreasing rate of growth. (3) changing direction of growth but not affecting the rate of growth.

5. When a growing plant is illuminated from one side only, cells are: (1) larger on the side nearest the light. (2) larger in the middle. (3) larger on the side farthest from the light. (4) all the same size

6. Which of the following tropisms is positive for the roots of a plant and negative for the stems? (1) Geotropism (2) Hydrotropism (3) Phototropism (4) Thermotropism

7. To which of the following stimuli does a worm respond positively? (1) gravity (2) light (3) touch (4) water

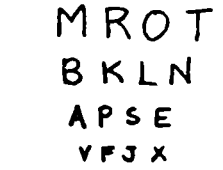
8. An earthworm's eye spots are located on its: (1) anterior dorsal surface. (2) anterior ventral surface. (3) lateral surface (4) posterior dorsal surface. (5) posterior ventral surface

9. An earthworm's mouth is located on its: (1) anterior end (2) dorsal surface (3) posterior end (4) ventral surface

10. A man wanted to test some new fertilizer. He used ordinary fertilizer on one field of corn and the new fertilizer on a second field of corn. The control is the: (1) corn (2) field with the new fertilizer (3) field with the old fertilizer (4) new fertilizer (5) old fertilizer

11. If a person was to guess whether a coin would come up heads or tails, in 12 flips he would most likely be right: (1) 2 times (2) 4 times (3) 6 times (4) 8 times (5) 10 times

Some scientists decide to find out whether grade 8 girls have better vision than grade 8 boys or vice versa. The test consists of reading 4 lines of letters each containing 4 letters. The letters in each row are made a little smaller than the row above. At the right is a sample of how these letters might appear. The following 4 questions refer to the experiments they use to solve the problem.



12. In choosing a sample of students to give the test to, they should choose (1) an 'average' grade 8 girl and an 'average' grade 8 boy. (2) 5 to 10 grade 8 girls and 5 to 10 grade 8 boys. (3) 30 to 40 grade 8 girls and 30 to 40 grade 8 boys. (4) grade 10 girls and boys because they can see better.

13. When giving the test it would be better to have one: (1) person give all the tests. (2) for the girls and one for the boys. (3) for each 3 or 4 students. (4) person for each student.

14. Which one of the following things is the least important factor to control?
(a) light in the room
(b) outside noise
(c) size of letters
(d) number of eyes used
(e) distance from chart
(f) whether it is a boy or a girl being tested

(1) a (2) b (3) c (4) d (5) e (6) f

15. If a person was just guessing at the letters in any row of 4, he would most likely get: (1) none right (2) one right (3) two right (4) three right (5) four right

A scientist had found a small organism which he was unable to identify as either plant or animal. The organism lived in pond water. He placed some in a dish which was half in darkness and half in light. After an hour he noticed most of the organisms had moved towards the light. He repeated the experiment 3 or 4 more times, with the same results.

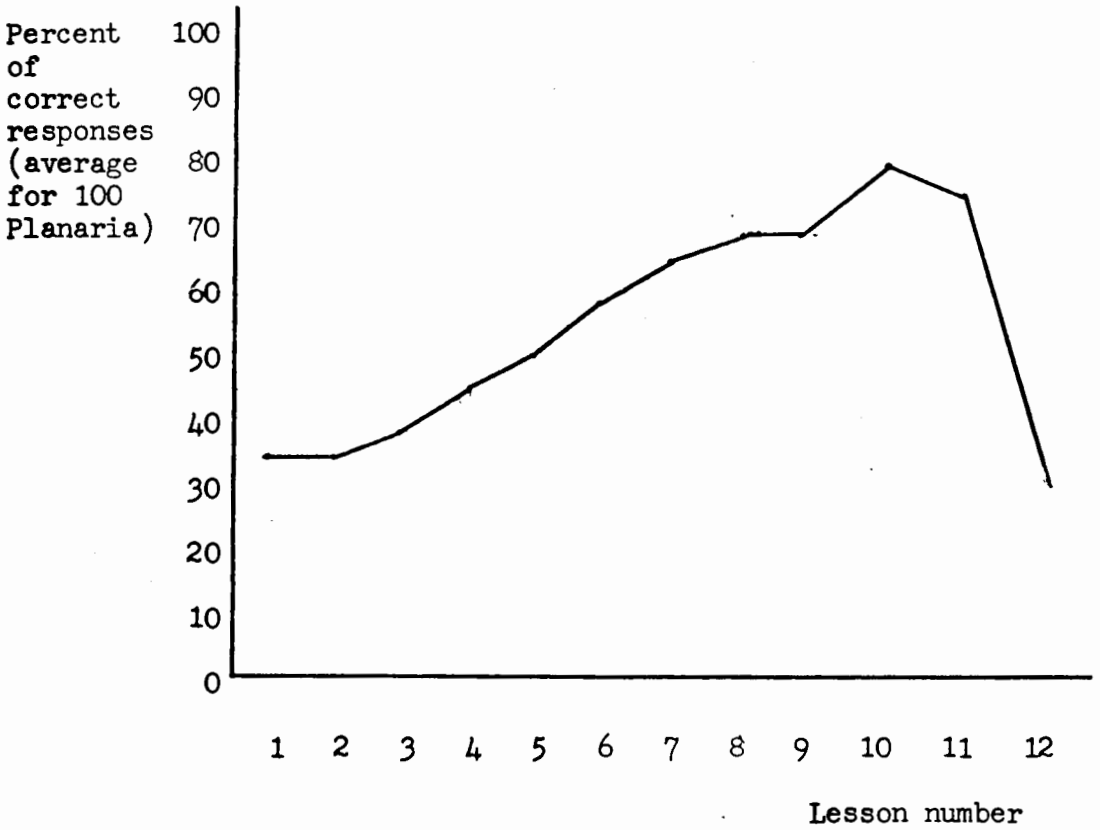
16. The organism's reaction to light was: (1) positive (2) negative (3) uncertain (4) phototaxis

17. The organism's reaction to water was probably: (1) positive (2) negative (3) uncertain (4) hydrotaxis

18. He repeated the experiment a number of times because: (1) the results of the first trial were not too clear. (2) He wished to be certain of the accuracy of his results. (3) his control group did not respond the first time. (4) his experimental group did not respond the first time.

A Hypothetical (Imaginary) Experiment

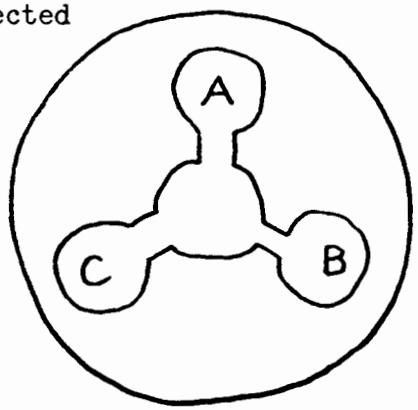
Number of correct responses after a certain number of lessons



No. of Learning Sessions

An experiment was done to find out whether Planaria (a small animal that lives on plants in ponds) were capable of learning to follow a simple maze. The maze was partially filled with stale water and looked like this:

The Planaria was put in the middle and was directed to go to a certain chamber (A, B or C). For half the Planaria, the correct chamber was indicated by having a green light in it while the other two chambers were lit by red light. The lights were reversed for the other Planaria so that red light indicated the correct chamber. During the lessons each time the Planaria went into the wrong chamber it was picked up and put back in the middle. Every time it went into the



correct chamber it was rewarded by a flow of nice, fresh water. After each lesson it was given a test with no rewards for the correct responses (going into the right chamber). The graph shows the average percentage of correct responses after each lesson. As a variation during the 11th lesson, the Planaria were punished by giving them an electric shock each time they made the wrong choice.

The following 7 questions refer to this experiment:

19. In order for this experiment to prove anything it must be assumed that: (1) planaria have no preference for one color over another. (2) planaria respond to red. (3) planaria respond to green. (4) someone else has been teaching these planaria.

20. The planaria were learning fastest at about the: (1) 2nd lesson (2) 5th lesson (3) 7th lesson (4) 10th lesson (5) 11th lesson

21. The planaria were learning least at about the: (1) 3rd lesson (2) 4th lesson (3) 6th lesson (4) 9th lesson

22. Punishment makes planaria: (1) forget (2) learn faster (3) learn slower (4) mad

23. If one were to catch a planaria in a stream and give it the maze test, one would expect it to get: (1) 0 - 10% right (2) 10%-20% right (3) 20% - 30% right (4) 30% - 40% right (5) 40% - 50% right

24. From the data from this experiment, which of the following is true?
(1) The effect of punishment lasts longer than that of reward.
(2) The effect of reward lasts longer than that of punishment.
(3) Reward and punishment have the same effect.
(4) Reward and punishment have opposite effects but they last the same time.

25. Which one of the following is the most important thing proved by this experiment? (1) it is possible to teach planaria. (2) planaria are not color-blind. (3) planaria do not like punishment. (4) planaria like fresh water.

Suppose we set up an experiment to see if water is necessary for germination. Four dishes of 10 corn seeds each are used.

Dish 1 - soaked seeds were placed on wet towelling
Dish 2 - soaked seeds were placed on dry towelling
Dish 3 - dry seeds were placed on dry towelling
Dish 4 - seeds were placed in $\frac{1}{2}$" of water

The seeds were set aside in the same place for several days and dishes 1 and 4 were kept from drying out. All were kept in the dark. Use this information to answer the next 4 questions.

26. The seeds which probably germinated first were in dish: (1) 1 (2) 2 (3) 3 (4) 4

27. The seeds which would probably not germinate were in dish: (1) 1 (2) 2 (3) 3 (4) 4

28. Which of the following was not a controlled factor: (1) using corn seeds. (2) experiment performed in the dark. (3) temperature (4) soaking corn seeds.

29. Dish 4 is used to show that: (1) the amount of moisture is important for germination. (2) seeds rot. (3) seeds need moisture to germinate. (4) a control is necessary.

30. What is the name of the taxis which is involved in the fact that Pacific Coast termites live only in damp wood: (1) phototaxis (2) hydrotaxis (3) thigmotaxis (4) geotaxis

31. What is the name of the taxis which is involved in the fact that fruit fly larvae (worm stage) move upwards in a tree from the place where the eggs are laid: (1) phototaxis (2) hydrotaxis (3) thigmotaxis (4) geotaxis

32. Which of the following is not a larvae of an arthropod: (1) grub (2) maggots (3) caterpillar (4) peripatus (5) meal worm

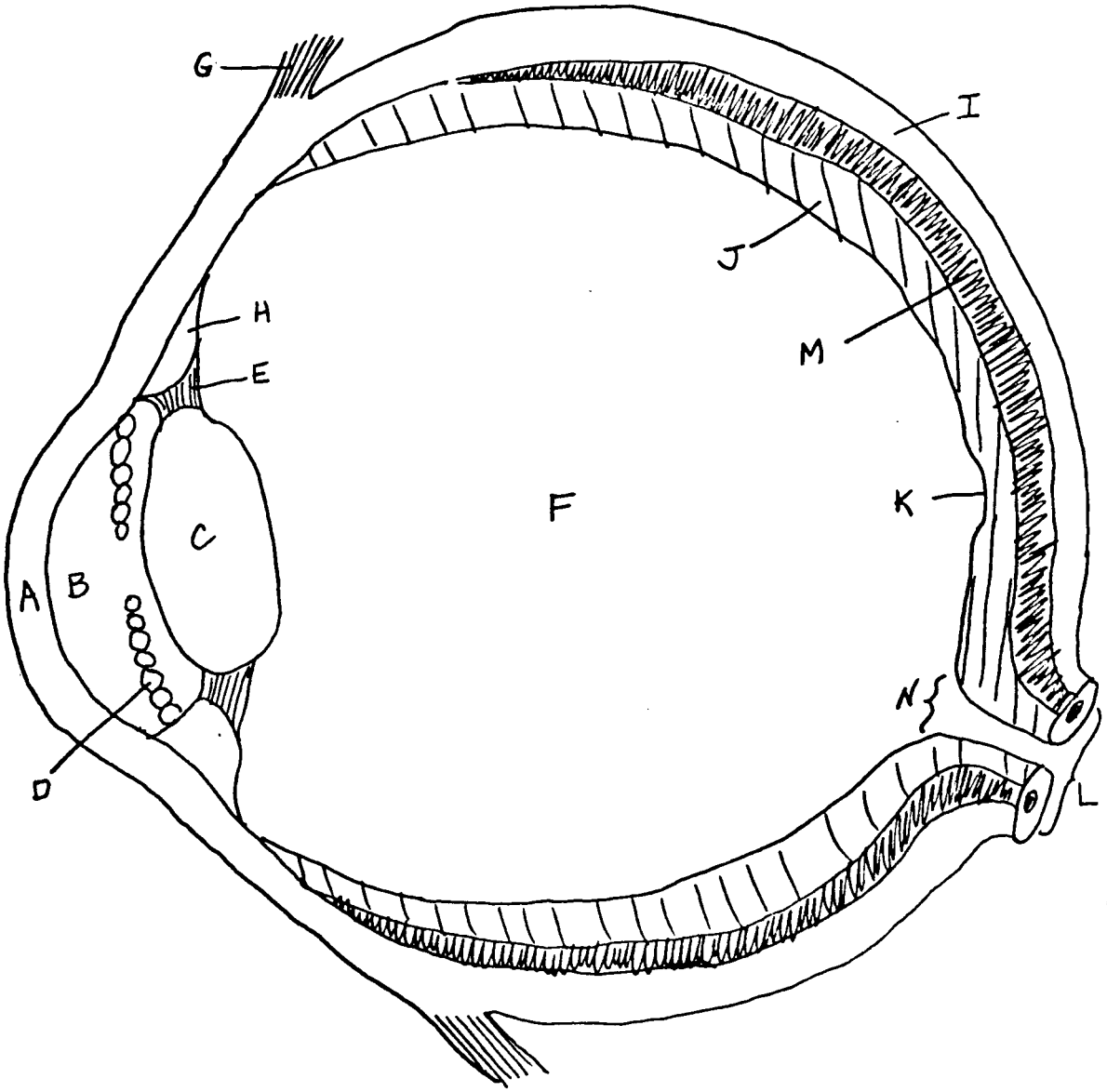
33. A sow bug belongs to the class of: (1) peripatus (2) arachnids (3) insects (4) crustaceans

34. An insect has _____ body sections: (1) 1 (2) 2 (3) 3 (4) 4

35. Which of the following is <u>not</u> an insect? (1) spider (2) dragonfly (3) termite (4) meal worm

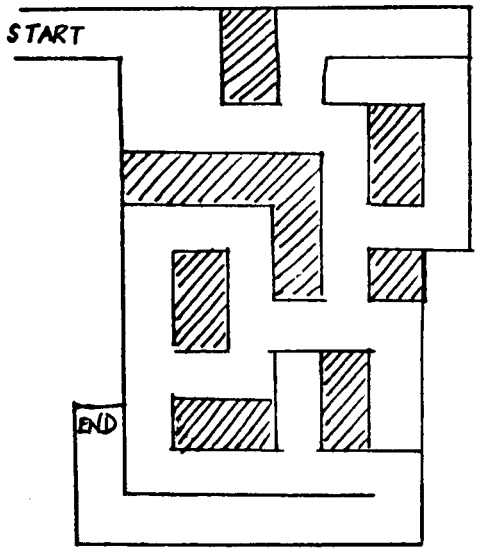
Using the sketch of the eye shown below - place the letter of the part opposite the names below.

36. aqueous humour _____
37. sclerotic coat _____
38. retina _____
39. choroid coat _____
40. iris _____
41. cornea _____
42. vitreous humour _____
43. lens _____
44. optic nerve _____
45. suspensory ligament _____



The following maze was used to experiment on learning in mice.

Use the information given to answer the following 5 questions.



Data for one mouse

Trial	Day	Time Taken
1	1	2 min. 5 sec.
2	2	2 " 3 "
3	3	1 " 20 "
4	6	2 " 10 "
5	7	1 " 35 "
6	8	1 " 2 "
7	9	0 " 40 "
8	11	1 " 20 "
9	12	0 " 55 "
10	13	0 " 40 "

46. How many 'dead-ends' are there in the maze for the mouse to take?
(1) 1 (2) 2 (3) 3 (4) 4

47. How many alternate routes (not including dead-ends) are there for the mouse to take from start to end? (1) 1 (2) 2 (3) 3 (4) 4

48. On which trial did the mouse learn fastest? (1) 3 (2) 4 (3) 8 (4) 9 (5) 10

49. The greatest amount of 'forgetting' took place between:
(1) trial 1 & 2 (2) trial 2 & 3 (3) trial 3 & 4 (4) trial 7 & 8

50. If a trial 11 was held on day 13 the most probable improvement in 'time taken' for the mouse would be about: (1) 5 sec.
(2) 20 sec. (3) 30 sec. (4) no improvement

Suppose we place some fresh, moist bread in some dishes and put a very little bit of mould into the centre of each dish. We then divide the dishes into 2 groups (A & B). We place Group A in a dark warm cupboard and Group B in a refrigerator. The dishes were covered so that the bread would not dry out. A week later the dishes were examined and Group A had much more mould than Group B.
Use this information to answer the next 2 questions:

51. Which of the following list was not a control in this experiment?
(1) temperature (2) amount of light (3) food for mould
(4) amount of moisture

52. If we are careful with the experiment, which one of the points below would be the most important conclusion to the mould experiment? (1) mould needs moisture to grow (2) mould grows on bread (3) new colonies of mould formed during the experiment
(4) mould needs warm temperatures to grow well

APPENDIX B

ACHIEVEMENT TEST

DESCRIPTIVE STATISTICS AND RELIABILITY DATA

APPENDIX B

Table 1

Achievement Test
Descriptive Statistics

range = 33	mean = 26.3
median = 26	variance = 46.7
mode = 20	standard dev. = 6.8

Table 2

Test/Retest Results
Over a Two Day Interval

Student	Test Score	Retest Score
1	24	23
2	18	24
3	34	33
4	20	22
5	29	28
6	32	32

Table 3

Split-Half Scores

Total Scores	Odd-Item Scores	Even-Item Scores
32	17	15
29	16	13
34	20	14
17	10	7
24	14	10
34	16	18
26	12	14
27	12	15
26	15	11
20	13	7
22	12	10
20	9	11
25	13	12
30	14	16
17	9	8
26	14	12

Table 4

A Comparison of Test Score Rank
and Teacher Rank of Achievement

Student	Test Rank	Teacher Rank
A	1.5	9
B	1.5	2
C	3	3
D	4	1
E	5	16
F	6	5
G	7.5	4
H	7.5	6.5
I	9.5	6.5
J	9.5	13
K	11	12
L	12	8
M	13	10.5
N	14	14
O	15.5	15
P	15.5	10.5

APPENDIX C

BREAKDOWN OF QUESTIONS ON ACHIEVEMENT BY LABORATORY

APPENDIX C

Breakdown of Questions on Achievement Test by Laboratory

Laboratory	Questions Related to Laboratory
1	3%
2	3
*3	14
*4	5
(1) 6	4
*7	13
(2) Earthworm Beh.	1
*8	8
9	5
*11	6
12	19
*13	19
	100%

* Indicates report that was marked for all groups.

(1) Laboratory 5 was not performed by the students as the live specimens were not available.

(2) This laboratory was from the source other than the text.

The laboratories marked (*) account for 65% of the questions on the achievement test.

APPENDIX D

SAMPLE MARKED LABORATORY REPORT

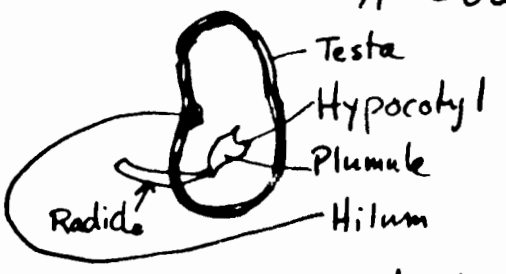
Name Eric Anzinger

Partner Burney Wood

2

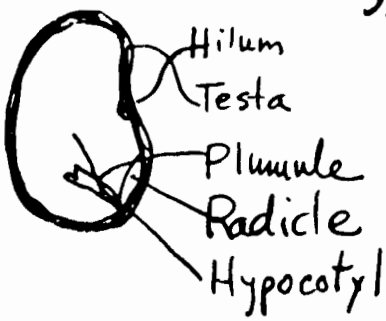
PROBLEM: Planting an Experimental Garden

A Soaked Bean



sketch magnification?

A dry Bean



Step 6

1a The soaked one is easier to ~~peal~~ peel

1B The soaking made the bean soggy and soft.

c The coating ~~protects~~ yes it from dirt and other things

D The soaked beans are sometimes bigger than the dry beans.

E The expanding of the outer cover.

Use full sentences please!

Step 8

2a The leaves are a gellowish white colour.

B I think ~~it~~ will turn into the leaves. ✓

C I think that the radicle will be the roots ✓

D I think that the hypocotyl will be the stems ✓

9 (a) hilum?

(b)

(c)

Are these really 4x actual size?