8-1970

# The Effect of Daily Quizzes on Hour Exam Performance

Katherine Janczarek

# THE EFFECT OF DAILY QUIZZES ON HOUR EXAM PERFORMANCE

by

Katherine M. Janczarek

A Thesis
Submitted to the
Faculty of the School of Graduate
Studies in partial fulfillment
of the
Degree of Master of Arts

Western Michigan University
Kalamazoo, Michigan
August, 1970

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to Dr. Richard Malott for his encouragement, advice and supervision during the course of this study and in the writing of this thesis. To the members of my committee, Dr. Howard Farris and Dr. Jack Michael, I express my gratitude for their criticism and help in preparing this report.

I thank Miss Linda Hilton for her time and skill in preparing this manuscript for submission to the Graduate School. To the faculty of the Psychology Department I express my appreciation for the intellectual training and financial aid of an Assistantship during my graduate career.

M-2509

JANCZAREK, Katherine Marie, 1948-
THE EFFECT OF DAILY QUIZZES ON HOUR EXAM PERFORMANCE.

Western Michigan University, M.A., 1970
Psychology, experimental

University Microfilms, Inc., Ann Arbor, Michigan

TABLE OF CONTENTS


| | Page |
|---|---|
| INDEX OF FIGURES | iv |
| INTRODUCTION | 1 |
| METHOD | 3 |
| Subjects | 3 |
| Procedure | 4 |
| RESULTS | 9 |
| DISCUSSION | 20 |
| CONCLUSION | 30 |
| REFERENCES | 31 |

INDEX OF FIGURES


| Figure | | Page |
|---|---|---|
| 1 | Distribution of combined exam scores for the Exam and Quiz groups . . . . . | 11 |
| 2 | Distribution of bonus point question scores for the Exam and Quiz groups . . . . . | 13 |
| 3 | Correlation between midterm exam scores and the number of quizzes missed for the Quiz group. . . | 15 |
| 4 | Correlation between final exam scores and the number of missed quizzes for the Quiz group . . | 17 |
| 5 | Correlation between midterm exam scores and the number of quizzes failed for the Quiz group. . . | 18 |
| 6 | Hypothetical curve of exam performance as a function of the amount of study time . . . . . | 23 |
| 7 | Hypothetical functions between amount of study time and exam performance when textual material is integrated and when it is discrete. Note: the heights of the curves do not represent superiority of the texts . . . . . | 26 |

Keller (1967 and 1968) introduced a behavioral approach to education stressing self-paced instruction, unit test perfection, optional motivational lectures, emphasis on textual material and the use of proctors as features of his new technology. Since the advent of these techniques, various applications of these methods have been made to the teaching of introductory psychology courses (Gallup, 1970; McMichael, 1968) with great success. Keller's methods have also been used to teach other areas including engineering (Koen, 1970) and statistics (Myers, 1969).

Michael (1969) has described the basic management of behavior in education at all levels in terms of the identification of goals, contingencies and consequences. Numerous studies have been conducted testing the application of these behavioral methods in elementary education (Levitt and Curtiss, 1969) as well as college level instruction. Sullivan (1969), Sheppard and McDermot (1970) and Marx (1970) investigated the effectiveness of personalized instruction with college students. Lloyd (1969), Malott and Svinicki (1969) and Malott and Palm (1970) have described the programming of course activities at an undergraduate level. Several investigators have developed new course materials which are organized into cohesive units with pertinent objectives to be used as study guides (Janczarek and Malott, 1969; Whaley and Malott, 1968; Mertens, Luker and Botuck, 1969; and Grosowsky, 1969). An undergraduate curriculum with courses based on a system of contingency management has been initiated with a high degree of success (Malott, Shook, Hartlep and Keenan, 1970).

Many of these studies have emphasized self-pacing which allows the student to progress through the course material at his own rate. According to Keller (1968), self-pacing is a very desirable feature of an instructional system. However, the task of programming material for a course which allows the self-pacing of over 1,000 students is difficult to administer. Because of this, an alternative method of daily quizzes over assigned reading material was employed (Malott and Svinicki, 1969). Their method also included remedial activities for those who found it necessary. Although the rate at which reading materials are covered is controlled by the instructor, there is a large amount of self-pacing involved. For example, students may adjust their study time according to their own needs or abilities. This allows the student to pass the quizzes on an individual basis. No penalty is imposed for passing a quiz in a remedial session.

The introductory psychology course at Western Michigan University has employed these teaching techniques for four years. Included in the course are daily reading assignments and daily quizzes. A large portion of the student's grade in this course is determined by his performance on these quizzes. Since the most essential feature of the course involves daily quizzes, this study investigated their effectiveness and influence on hour exam performance.

## METHOD

### Subjects

During the Winter semester of 1970, students enrolled in Psychology I: Introduction at Western Michigan University volunteered to participated in this study. Any student who volunteered would be exempt from certain daily activities which included student centered discussion groups designed as a framework to analyze and criticize articles on the basis of the attitudes of science and experiments performed in the laboratory section of the course.

Since it is often difficult to conduct educational research within a course without considerable student complaints, the students were given a choice as to whether they would prefer the course as it was normally offered or as in this study. However, students could not elect to change groups or return to the normal course once they had been assigned to an experimental group.

Students were given an explanation of the experiment and then asked to apply for participation. From the list of volunteers, students in each section of the course were randomly assigned to one of two experimental groups. A total of 185 students volunteered and participated in the experiment with 92 students in one group, designated Quiz group; and 93 in the other, designated Exam group. Twenty students in this study dropped the course before the end of the term; eight in the Exam and twelve in the Quiz group.

## Procedure

The manner in which the course text material was presented was the main independent variable, while the main dependent variable was performance on mid-term and final exams.

The text material was taken from two main sources: Analysis of Behavior, a programmed text by Holland and Skinner (1961) and Elementary Principles of Behavior by Whaley and Malott (1968). Each student was given a syllabus which presented the material in daily one-hour reading assignments. Each student also had a book, The Big, New, Mother, Mind-Boggling, Behavior-Expander by Janczarek and Malott (1969) which contained objectives or study questions for each assignment with a list of terms and principles. Any supplementary or clarifying materials developed throughout the term were distributed to all students.

## Grading

The Exam group was required to attend class twice during the semester to take a mid-term and final exam. The combined score on the tests was used as the measure of performance on the exams. Grades for the course were assigned to the students in this group on the basis of a curve similar to curves used by other freshman courses at the university. A grade curve for a large general studies humanities course in the Winter of 1969 was used as the model. According to this curve, 15.1% of the students received an "A", 47.3% received "B's, 29.8% received "C"s, 4.5% received "D"s and 1.2% of

the students received "E"s in the course.

The Quiz group was required to attend class daily, Monday through Thursday, to take a two point quiz over the daily reading assignment. These daily quizzes consisted of fill-in-the-blank, multiple-choice, or short answer type questions. If a student failed to pass this quiz given in class by answering both questions correctly, remedial opportunities were available throughout the following day. Make-up quizzes were of a conceptual nature. Each quiz was composed of a short description of some behavior followed by five or six multiple-choice questions which asked the student to identify and analyze specific concepts illustrated in the description. In order to receive passing credit for any missed quiz, the student must have answered all questions on a remedial quiz correctly.

For purposes of clarity, any quiz given in class that a student did not pass will be called a missed quiz. Any quiz that a student missed and did not remediate will be termed a failed quiz.

The Quiz group also took the mid-term and final exams. Grading for this group was based on quiz and exam performance. The quiz grade was determined by the number of daily quizzes passed, either in class when the first quiz was given or at the remedial session over that reading assignment. Letter grades were assigned for quiz performance on the basis of the number of quizzes failed. The grade scale for the quiz activities was as follows:

| | | |
|---|---|---|
| 0-2 | quizzes failed ------- | A |
| 3-4 | quizzes failed ------- | B |
| 5-8 | quizzes failed ------- | C |
| 9-12 | quizzes failed ------- | D |
| more than 12 | quizzes failed ------- | E |

The exam grade for the Quiz group was determined on an absolute standard. The percentile scores resulting from the curve used to compute the course grade for the Exam group were used as the minimum point values necessary for a given grade on the exams. For example, the top 15.1% of the students in the Exam group had combined scores between 74 and 90 which then defined the "A" range for the Quiz group. For the Exam group, 47.3% had scores between 60 and 73, the "B" range; 29.8% had scores between 47 and 59, the "C" range; 4.5% scored between 38 and 46, the "D" range; while 1.2 scored lower than 38 on the exams. These scores served as the absolute standards to which exam scores for the Quiz group were compared.

Course grades for the Quiz group were based on the lower of the grades obtained in the two activities. However, in an attempt to alleviate some of the aversive aspects of the exams, the exam grade could lower the student's grade by only one letter. In other words, if a student maintained an "A" level of performance on the daily quizzes but received a "C" for a grade on the exams, the course grade for this student would be lowered to a "B" rather than a "C".

Exams

Two, fifty question, multiple-choice tests were given as the mid-term and final exams. Questions were chosen, for the most part, directly from the questions available for daily quzzing. In cases where fill-in-the-blank or short answer questions were picked, the question was transformed into a multiple-choice format.

Four forms of each test were compiled in order to reduce the likelihood of collusion between students during the testing sessions. On the mid-term exam, Forms A and B contained questions which were different from Forms C and D. The questions on Forms A and B, while the same, were arranged in different order. The questions on Forms C and D were also arranged differently. A similar procedure was used for the final exam.

The exams were administered during the regularly scheduled Friday class period which met in a large lecture hall. During each testing session, all forms of the exams were given. Answers were marked on IBM mark-sense sheets and computer graded by Testing Services at the university.

The final exam contained ten bonus questions of a True-False nature. Students were asked to correctly identify whether a particular case study was presented in the Whaley and Malott text. For each question correct, one point was added to the student's score after a grade for the exam performance was determined, thus increasing the probability of a higher exam grade. Therefore, the actual course grades received were somewhat higher than would have

been obtained according to the original humanities grade distribution.

Correlations

Remediation effectiveness was measured by computing correlations between the number of quizzes missed in class for the Quiz group and the exam scores obtained by this group. The number of quizzes failed (i.e. missed and not remediated) were also correlated with exam scores for this group.

## RESULTS

The distribution of combined exam scores for the Quiz and Exam groups appears in Fig. 1. The mean score out of a possible of 90 points (mid-term plus final, minus bonus points) was 62.7 (69%) for the Exam group and 66.3 (73.6%) for the Quiz group. This difference between means results in a $t$-value of 3.36. For 157 degrees of freedom, a $t$-value of 2.58 is significant at the .01 level of significance.

The distribution of bonus point scores for both groups is presented in Fig. 2. The mean score out of a possible of 10 points for the Exam group was 6.35 (63.5%) while the mean score for the Quiz group was 7.30 (73%). This difference between means results in a $t$-value of 4.48. For 157 degrees of freedom, a $t$-value of 2.58 is significant at the .01 level of significance.

Correlations between the number of quizzes missed in class and the mid-term and final exam scores for the Quiz group are presented in Figs. 3 and 4. The correlation coefficients computed for these data were shown to be significant different from zero ($r$ = -.73 and $r$ = -.35 respectively) at the .01 level of significance.

Correlational data between the number of quizzes failed and scores on the mid-term exam for the Quiz group are presented in Fig. 5.

Figure 1. Distribution of combined exam scores for the Exam and Quiz groups.

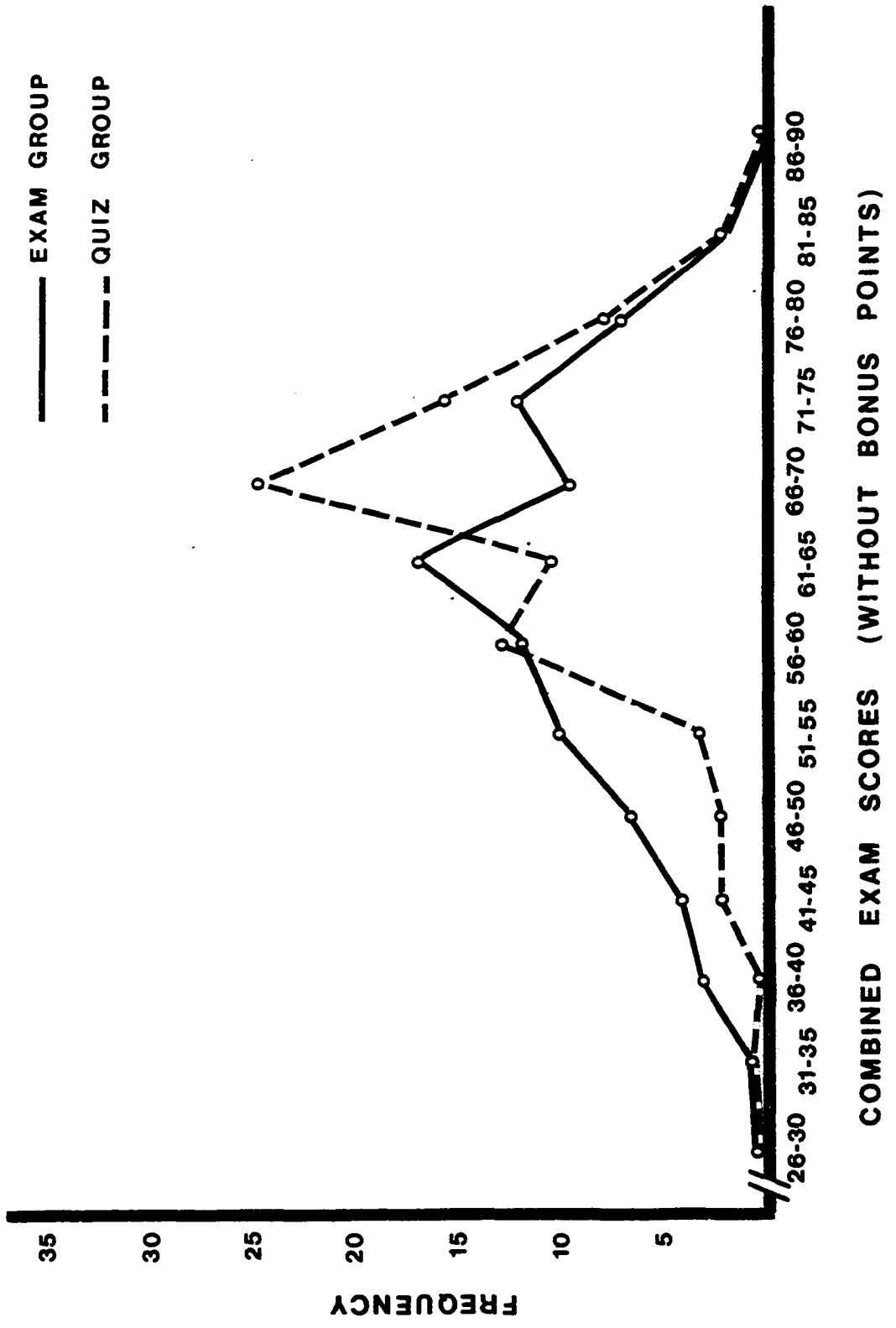


Figure 2. Distribution of bonus point question scores for the Exam and Quiz groups.

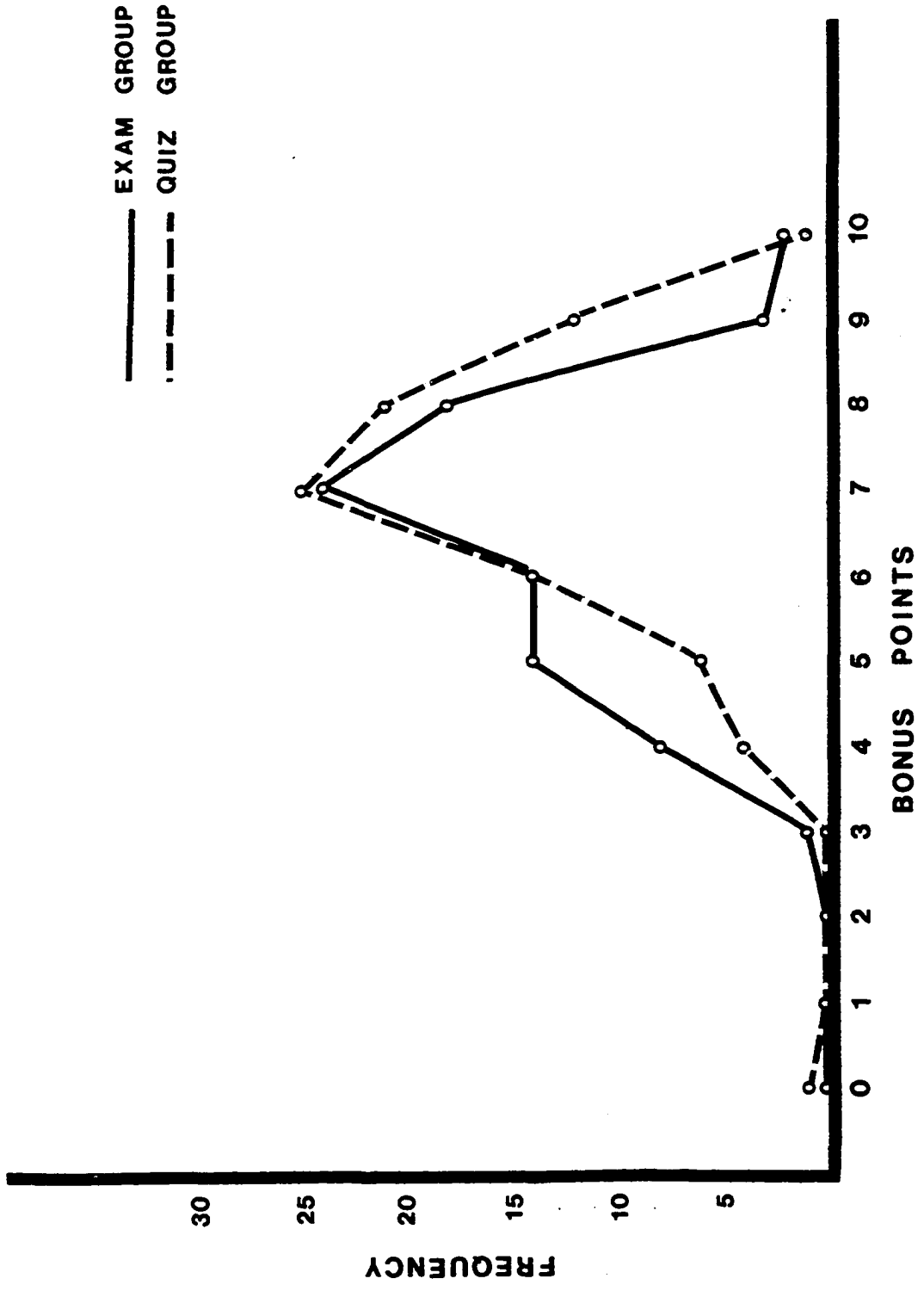


Figure 3. Correlation between midterm exam scores and the number of quizzes missed for the Quiz group.

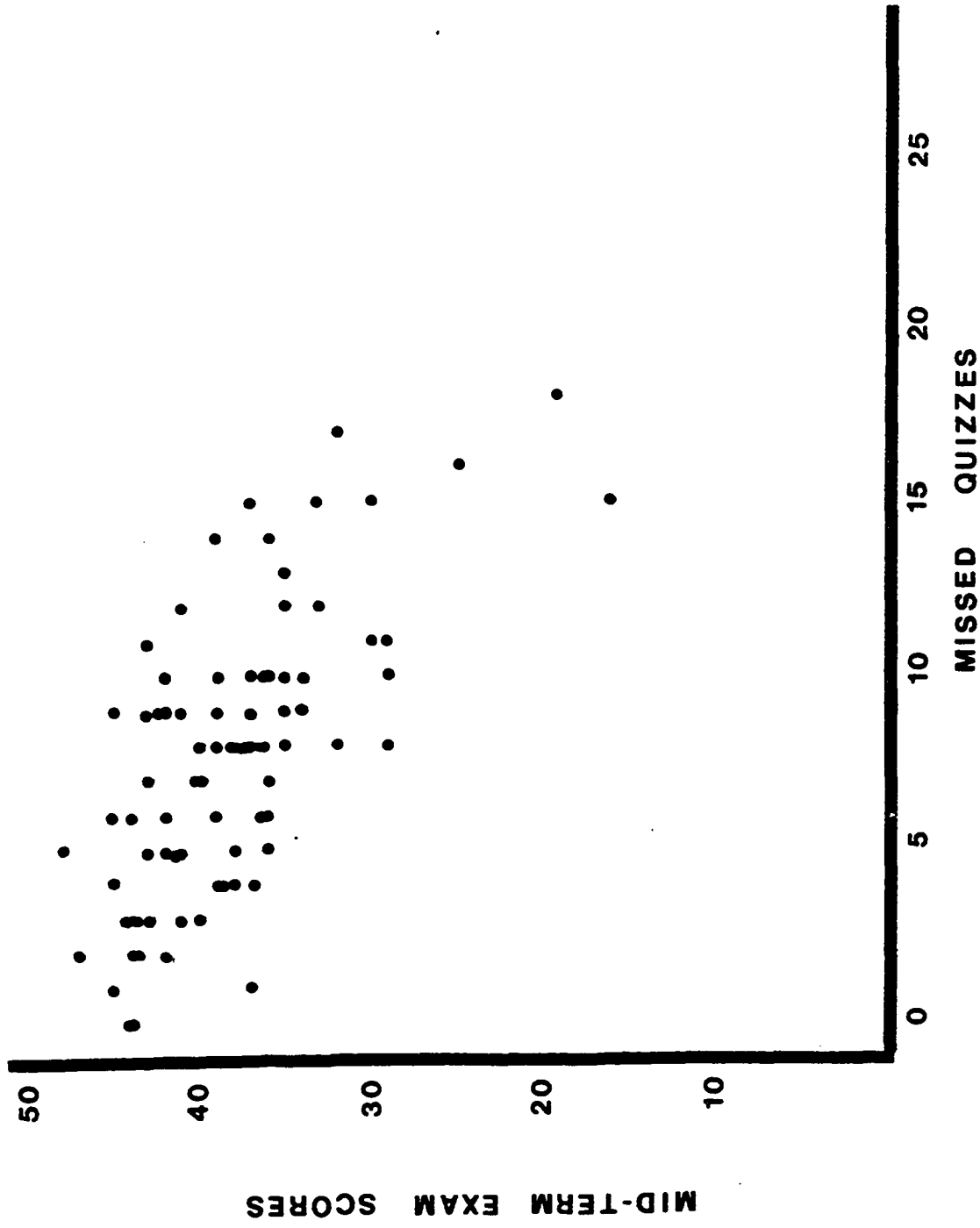


Figure 4. Correlation between final exam scores and the number of missed quizzes for the Quiz group.

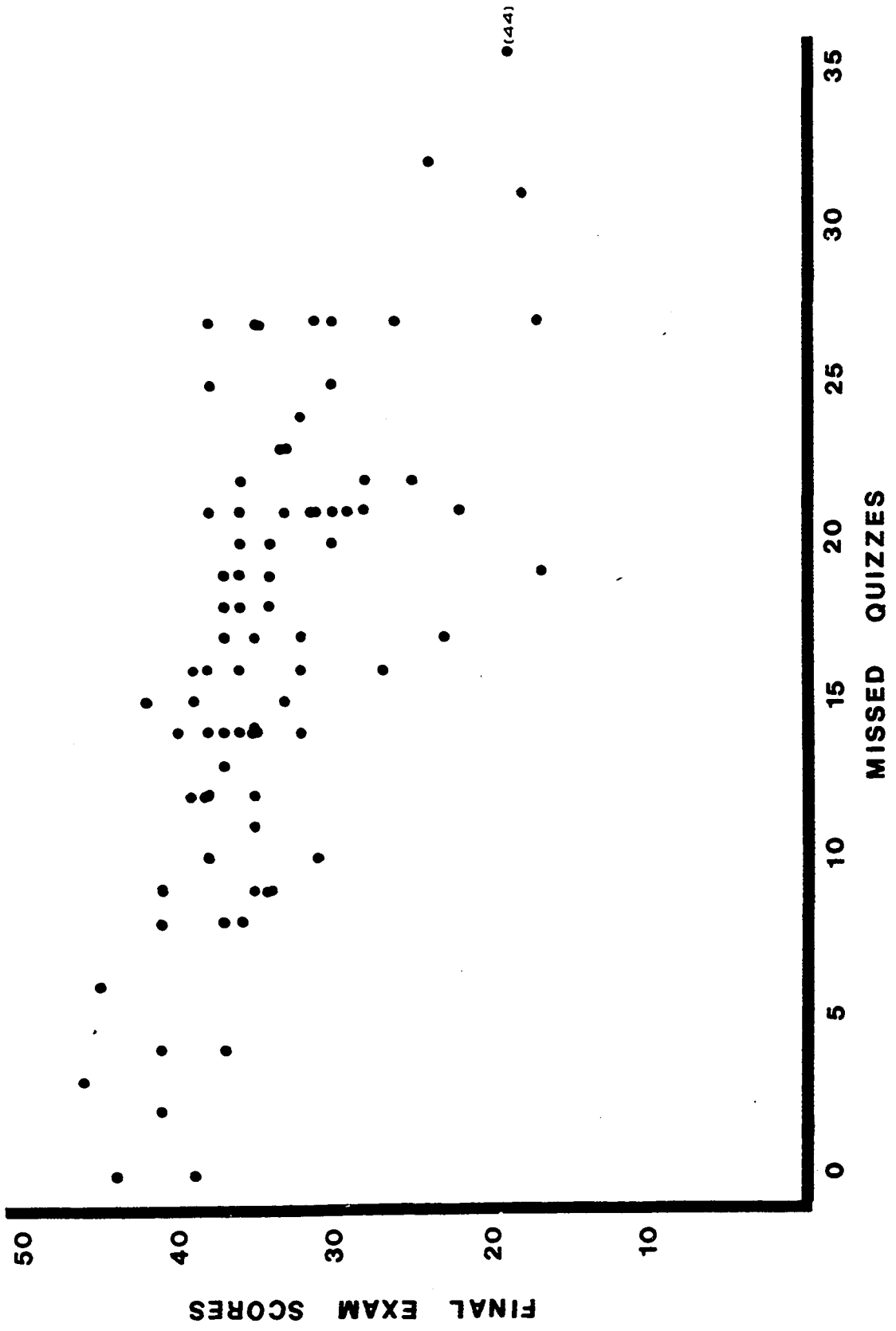
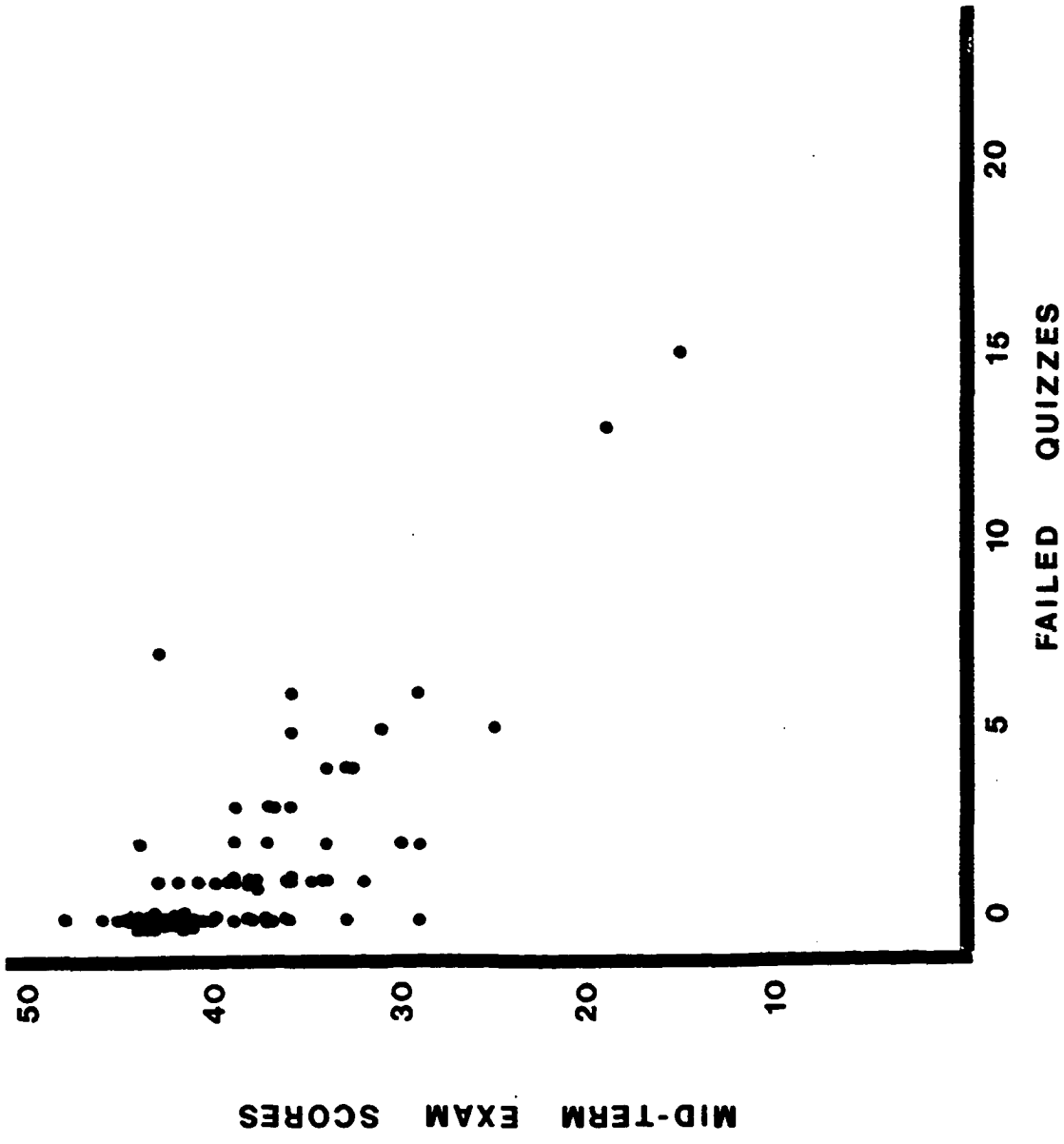


Figure 5. Correlation between midterm exam scores and the number of quizzes failed for the Quiz group.



A test of significance showed the correlation coefficient to be significantly different from zero at the .01 level ($r = -.367$). The correlation between the number of failed quizzes and final exam scores was so small that it could occur by chance with a probability greater than .05 when the population correlation is zero.

## DISCUSSION

The data presented here demonstrate that the use of daily quizzing techniques produced statistically signficant differences between the two groups of students. However, since the actual point differences between the means of the Exam and Quiz groups was relatively small, only 4%, the practical signficance of the results was less obvious. The instructors of large introductory college courses might be reluctant to invest the resources needed to restructure their present teaching methods to accomodate the use of daily quizzes. Therefore, an analysis of the parameters affecting the size of the point difference will be considered.

Two behavioral relations might be relevant to exam performance. The quiz schedule employed can affect the amount of study behavior; and, in turn, the amount of study behavior can affect the degree of performance on exams.

Bostow, et al. (1970) reported that the regularity of study behavior was a function of the quiz schedule employed. When students were given daily quizzes, they read the assigned material daily with little or no variation. When the quiz schedule was made more intermittent, study behavior became variable and showed a scalloped response pattern. However, no marked affect in the total amount of time spend in study was observed in this investigation. Also, data were not presented concerning the proportion of reading material actually studied under the different quiz schedules. Casual obser-

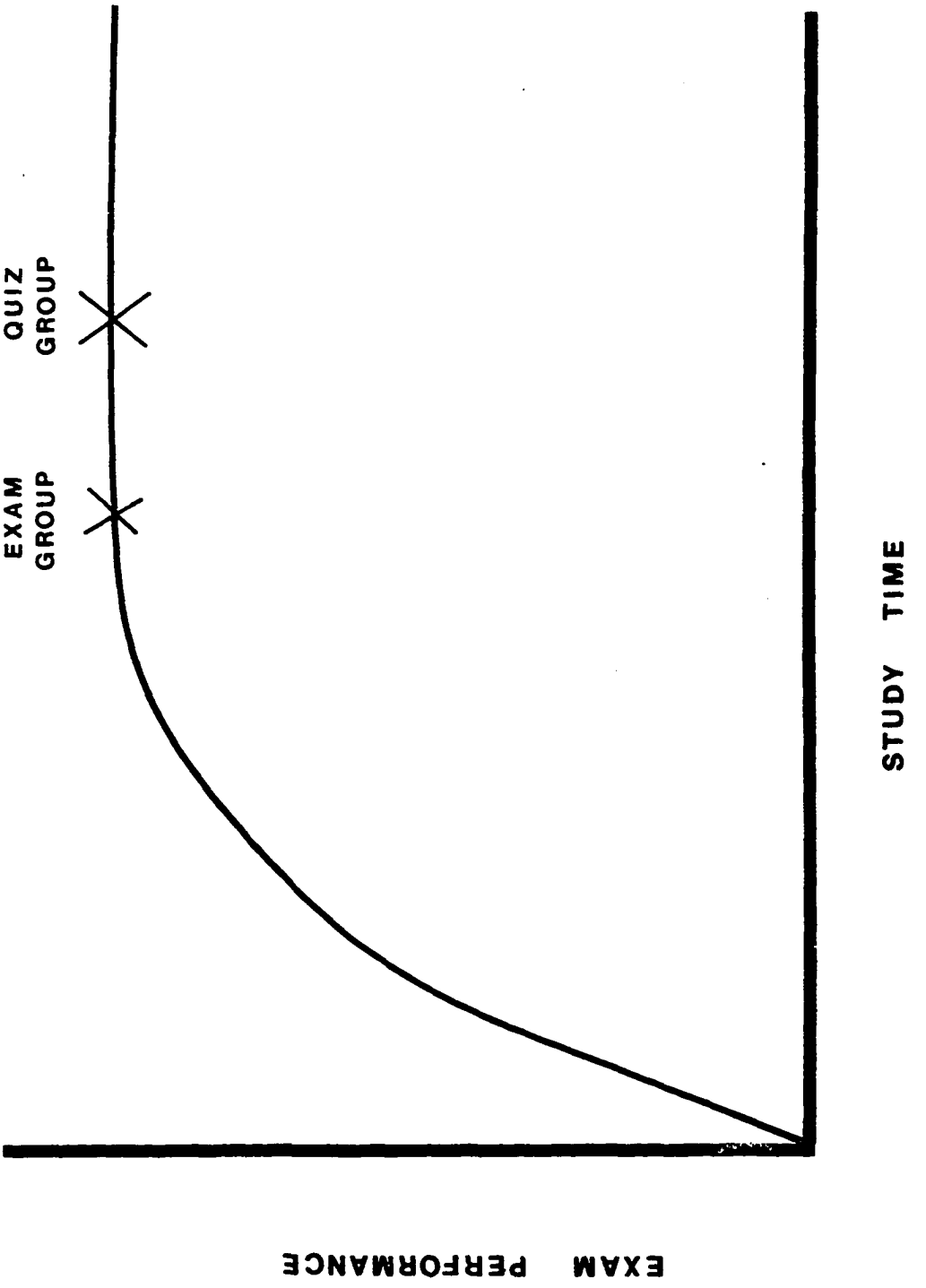
vation of student study behavior seems to indicate that with large amounts of material, (50-60 hrs. of reading), daily quiz procedures produce reading of approximately 90% of the assigned material; but with traditional testing procedures, students seem to read much less. Of course, systematic verification of this observation is needed.

Even if the quiz schedule does affect the amount of study behavior, the question still remains: does the amount of study affect the performance attained on an exam? The exam score is most likely a curvilinear function of the amount of study time. The student may reach a point of diminishing returns where additional studying does not greatly increase his mastery of the material to be learned (see Fig. 6).

If the Quiz and Exam groups studied approximately the same amount of time in preparation for the exam, a large difference between their scores would not be expected. Even if the two groups spent different amounts of time in study but were near the asymptote of the curve, large differences between the scores would not be obtained.

The shape of the function between study time and exam performance would be determined by the nature of the reading material and the exams themselves. In this study, both main texts were arranged so that each succeeding chapter or programmed set was intimately related to the preceding unit. Each assignment, while presenting a few new concepts, was also designed to produce an increased mastery of the subtle concepts previously presented. Traditional psychology texts often involve discrete chapters of information which include an abundance of facts with little integration or redundancy.

Figure 6. Hypothetical curve of exam performance as a function of the amount of study time.



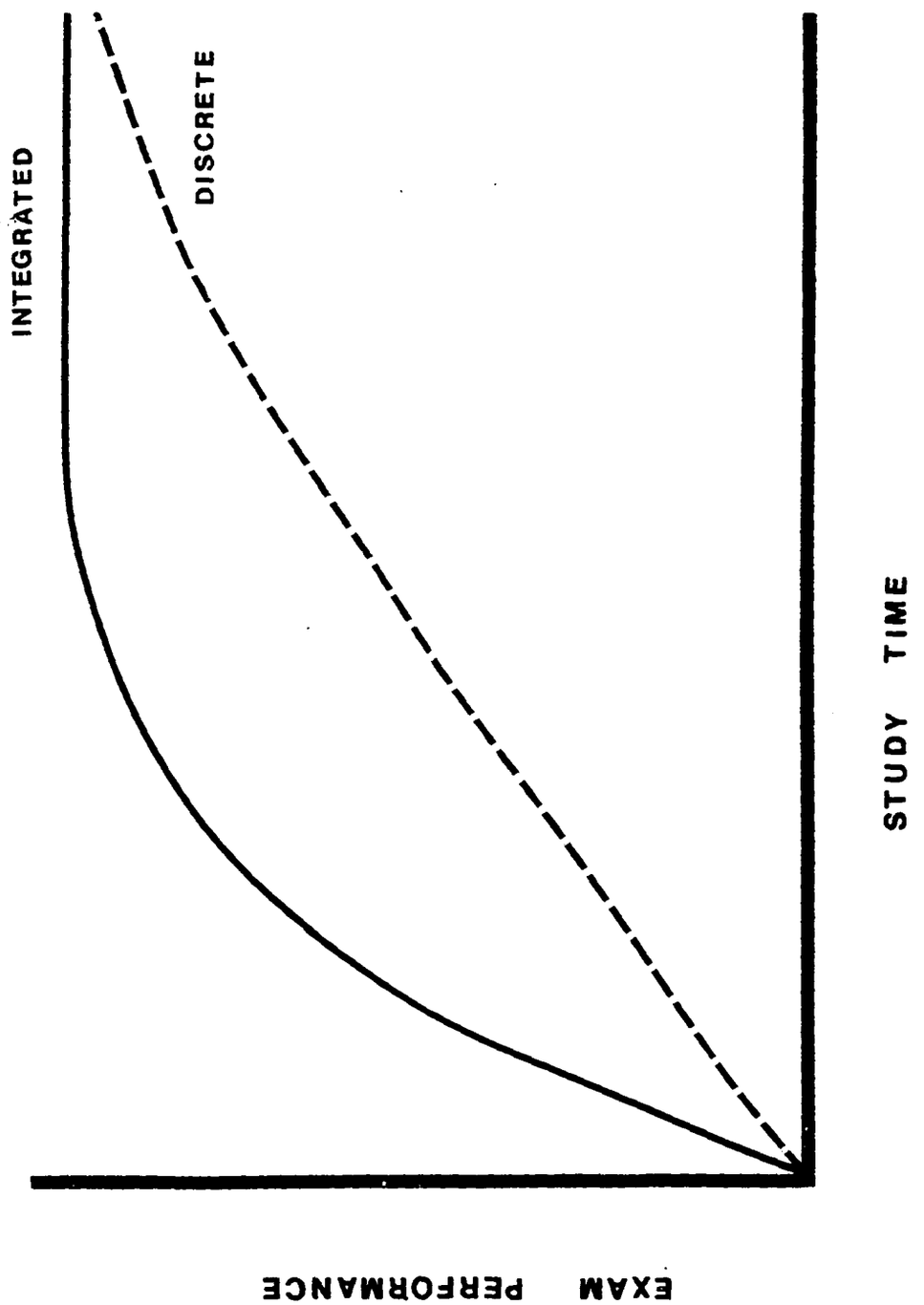
The difference in the shape of the performance curves produced by these two types of material is demonstrated in Fig. 7. In the case of the redundant and integrated material, the asymptote might be more quickly reached. With the traditional material, the function should become more linear: increases in study time result in increases in exam performance.

Since the course material and exams stressed terms, principles and the recognition of certain behavioral processes in novel situations, students in the Exam group could have reached a relatively high level of mastery simply from intensive study of the terms, principles and chapter summaries available to them. Students in the Quiz group would have studied the same material but would also have read the entire material assigned. The additional time spent in studying the material would increase their exam score very little when the relationship is curvilinear.[1]

---

[1]Students during the Spring of 1970, were given an exam after they had read only the chapter summaries of the Whaley and Malott text. The mean score on this exam was 60%. The same exam was administered after the students had read the body of the text. The mean score on this second exam was 73.4%, an increase of 12.4%. Even though this exam was specifically designed to test material available in the chapter summaries, the results indicate that reading the entire text material increases performance on this exam. Another exam (the final exam given during the Winter of 1970) was also administered after the students had read the entire Whaley and Malott text. The mean score on this exam was 60%. The Holland and Skinner programmed text was then assigned. The final exam was given again after this text was read. The mean score on this exam was 68%, an increase of only 8%. These data indicate that the function between the amount of study time and exam performance in our course is, in fact, quite curvilinear, as would be predicted on the basis of the content redundancy.

Figure 7. Hypothetical functions between amount of study time and exam performance when textual material is integrated and when it is discrete. Note: the heights of the curves do not represent superiority of the texts.



There is also a way in which the nature of the exams themselves can determine the shape of the curve. The more the exam performance is dependent upon the student's repertoire established prior to the course, the less the performance would be affected by course study time. Furthermore, if the exams were relatively simple, differences in study time would not greatly affect the performance of the student.

The average score on the exams for the Quiz and Exam groups was 71% which seems to indicate an appropriate level of difficulty. An analysis of the questions themselves showed them to be relevant to the information being taught. Questions were also carefully screened in order to eliminate those which were most ambiguous or confusing.

The inclusion of the bonus point questions on the final exam was an attempt to evaluate the extent to which students in each group could identify case studies taken from the course texts. It was felt that these detailed questions would serve as a more sensitive index of mastery of the material. The results indicated that this was the case. There was a 4% difference in mean scores on the exams without the bonus points. However, a difference of 10% between mean scores on the bonus questions was observed. Students in the Quiz group were able to identify a significantly greater number of the details from the case studies than the students in the Exam group.

The bonus point questions were also included in the final exam to measure the difference in performance between the two groups on questions which were more typical of factual questions often employed

in traditional courses. If differences in scores on the mid-term and other questions on the final exam were due to idiosyncrasies of the text or testing material, the bonus questions would not be subject to these influences. The difference between the exam score means for the two groups could have been due to an artifact. Students in the Quiz group might have scored higher on the regular exam questions due to pre-test familiarization with the questions. Since quiz and exam questions were taken from the same pool, the exams could have included questions which the Quiz group were given as quiz questions in class. However, the differences between mean scores on the bonus point questions could not be confounded by this variable.

Another problem examined during this study was the effectiveness of remediation. Correlations between mid-term and final exam performance and the number of quizzes missed in class were shown to be statistically significant. A correlation between mid-term exam performance and the number of quizzes failed was also shown to be statistically significant. However, a correlation between final exam performance and the number of quizzes failed was not significant. If the present remediation procedures are effective then there should be no quiz failures and no correlation between exam performance and the number of quizzes missed.

One would not expect the remediation procedures to be effective if students never availed themselves of the remedial sessions. Data indicate, however, that of the 1,351 quizzes missed throughout the course of the semester only 165 quizzes were failed. Remediation

reduced the number of failures to the point where the number of quiz failures did not correlate with the final exam performance. However, a correlation still remained between exam performance and the number of missed quizzes. On the basis of these data, it can be concluded that remediation is more successful in reducing the number of quiz failures than attenuating the effects of missed quizzes on the exam performance. It is not clear what the exam performance would have been if no remediation were available. Further studies exploring the various effects of remediation of poorly mastered subject matter should be carried out in the future.

## CONCLUSION

This study has shown that a statistically significant difference in mean scores on hour exams can be obtained through the use of daily quizzing techniques. The quiz schedule was assumed to directly influence the amount of study time and thereby indirectly influence the examination performance. The nature of the text material and testing materials were analyzed in terms of their affect on mastery of the course content as indicated by performance on the mid-term and final exams.

REFERENCES

Bostow, D. E., Mawhinney, V. T., Laws, D. R., and Blumenfeld, G. J., An analysis of student studying behavior as a function of two schedules of testing. Paper read at the annual meeting of the American Educational Research Association., Minneapolis, 1970.

Corey, J. R., McMichael, J. G., and Tremont, P. J., Longterm effects of personalized instruction in an introductory psychology course. Paper read at the 41st Annual Meeting of the Eastern Psychological Association., Atlantic City, 1970.

Gallup, H. F., Individualized instruction in an introductory psychology course. Paper read at the 41st Annual Meeting of the Eastern Psychological Association, 1970.

Grosowsky, H., Materials prepared for general studies course, Man's Contemporary Environment. University of Southern Illinois, 1968.

Janczarek, Katherine M. and Malott, R. W., The Big-New-Mother, Mind-Boggling, Behavior-Expander. Kalamazoo: Behaviordelia, 1969.

Keller, F. S., Neglected rewards in the educational process. Paper read at the 23rd Annual Meeting of the American Conference of Academic Deans., Los Angeles, 1967.

Keller, F. S., "Good-bye, teacher...". The Journal of the Applied Analysis of Behavior, 1969, 1, 79-89.

Koen, B. V., Self-paced instruction for engineering students. Engineering Education, 1970, vol. 60, no. 7, 735-736.

Lovitt, T. C. and Curtiss, Karen A., Academic response rate as a function of teacher- and self-paced contingencies. The Journal of the Applied Analysis of Behavior, 1969, 2, 49-53.

Lloyd, D. E. and Knutzen, Norma J., A self-paced programmed undergraduate course in the experimental analysis of behavior. The Journal of the Applied Analysis of Behavior, 1969, 2, 125-133.

Malott, R. W. and Palm, J. M., Contingency management revisited. Paper read at the Michigan Academy of Science, Detroit, 1970.

Malott, R. W., Shook, G., Hartlep, Particia and Keenen, M., The student-centered education project. Paper read at the Michigan Academy of Science., Detroit, 1970.

Malott, R. W. and Svinicki, J. G., Contingency management in an introductory psychology course for one thousand students. The Psychological Record, 1969, 19, 545-556.

Marx, R., General psychology at the college level for high risk students. Unpublished manuscript, 1970.

McMichael, J. S. and Corey, J. R., Contingency management in an introductory psychology course produces better learning. The Journal of the Applied Analysis of Behavior, 1969, 2, 79-83.

Mertens, G., Luker, A., and Botuck, C., Behavioral Science Behaviorally Taught. Minneapolis: Burgess Publishing Co., 1969.

Michael, J., Management of behavioral consequences in education. Revisita Interamericana de Psicologia, 1969, 205-252.

Myers, W. A., Operant leaning principles applied to teaching introductory statistics. Unpublished manuscript, 1969.

Sheppard, W. C. and MacDermot, H. G., Design and evaluation of a programmed course in introductory psychology. The Journal of the Applied Analysis of Behavior, 1970, 2, 28-34.