8523096

Parker, Dennis Keith

THE EFFECT OF TEST-LIKE EVENTS DURING MATHEMATICS LECTURE

The University of Oklahoma PH.D. 1985

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE EFFECT OF TEST-LIKE EVENTS DURING MATHEMATICS LECTURE

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

By

DENNIS KEITH PARKER

Norman, Oklahoma

1985

THE EFFECT OF TEST-LIKE EVENTS

DURING MATHEMATICS LECTURE

Approved By

Thomas J. Hill

Dillon J. Rogers

George W. Letchworth

Gerald Kidd

Curtis C. McKnight

Larry E. Toothaker

ACKNOWLEDGEMENTS

I wish to thank, with the greatest sincerity, all those who contributed to this study. I am especially appreciative of Dr. Thomas J. Hill, Chair of the dissertation committee, particularly for the encouragement and reinforcement he gave me during my years of graduate study. I have the utmost respect for him, both personally and professionally.

I wish to express gratitude to Dr. Gerald D. Kidd, Dr. George A. Letchworth, Dr. Curtis C. McKnight, Dr. Tillman J. Ragan, Dr. Patricia L. Smith, and Dr. Larry E. Toothaker for their time and helpful suggestions. Dr. Stanley B. Eliason, Associate Chair of the mathematics department, also deserves recognition. Without his cooperation, this study would not have been possible.

Finally, I would like to thank Kathleen Hendrix and my parents, James and Joy Richards, for their love, support, and understanding.

TABLE OF CONTENTS

LIST OF TABLES

TABLE		Page

THE EFFECT OF TEST-LIKE EVENTS DURING MATHEMATICS LECTURE

CHAPTER I

INTRODUCTION

Background of the Study

An increase in the activation of the cognitive processes of students during instruction is a key element of effective instruction as long as the activation is focused on task. A disparity exists, however, in that the majority of college class instruction is in a lecture format with students in a passive role. This is not conducive to much activation of a student's cognitive processes. Osterman (1982) cited studies documenting the fact that the vast majority of classtime is spent in the traditional one-way form of communication with little active involvement or feedback for the student. In particular, the most typical method used by college mathematics instructors is lecturing in conjunction with writing on the chalkboard or an overhead projector. The students respond by taking notes of what has been written. While there may be other purposes of a lecture, such as promoting interest in the content, if the main function is engendering learning, it is questionable whether this method is effective in stimulating the processes required for the learning of mathematics.

With this lecture method being so predominant in college mathematics classes, the nature of student attention during lecture is of primary concern. Several studies investigated the nature of student attention during lecture. Lloyd (1967) measured student attention by the quality of notetaking during lecture and found that effective student attention is optimal after about seventeen minutes, then declines to about fifty percent of that peak after fifty minutes. There is a rise in the level of attention during the last five minutes of lecture. Other studies measured student attention during lecture by testing for immediate recall and long-term retention of the content. Trenaman's (Berndt, 1971) findings, using both recall and retention of the lecture material as evidence of student attention, are consistent with the results of the Lloyd study; after fifteen minutes, the lecture fails to communicate effectively. Cameron and Giuntoli (1972), in studying undergraduate education students, found that only forty-six percent of a class is paying good attention at any given moment.

To address this dilemma, educational researchers have concerned themselves with how to integrate active involvement with instruction and have investigated the conditions for which active involvement facilitates learning. Based on research with programmed instruction, the interspersing of test-like events (questions or problems requiring a response from each individual) during lecture appears to be a possible solution (Anderson, 1970). There are a number of studies dealing with the effects of test-like events (commonly referred to as adjunct questions) during instruction. One of the more interesting effects relates to the type of learning that is measured. Rothkopf (1966, 1970) opened up an entire area of research by investigating the general facilitating effects of adjunct questions. While it is no surprise that adjunct questions facilitate achievement on a criterion test consisting of questions directly related to the adjunct questions, Rothkopf was primarily interested in the effect on a criterion test consisting of questions only tangentially related to the adjunct questions. Hence, these learning outcomes are commonly referred to as "indirect effects," "general effects," or "incidental learning." Rothkopf and other researchers (e.g., Boker, 1974; Bruning, 1968) provided support for the notion that adjunct questions may facilitate incidental learning. However, the effects found are often quite small and sometimes not statistically significant (Ladas, 1973). Dayton (1977) reviewed the variables that researchers studied to determine optimal conditions of test-like events. These variables include question type, age of student, immediate vs. delayed testing, knowledge of correct response, motivation, student ability level, and question placement. Dayton concluded that various combinations of these variables may facilitate learning.

One variable that seems to be crucial in the facilitation of incidental learning is question placement. Questions given prior to instruction (pre-questions) may facilitate direct learning but seem to inhibit incidental learning. On the other hand, questions given following instruction (post-questions) may facilitate both direct and incidental learning (Rothkopf, 1966).

A few studies dealt with test-like events used in mathematics text. Mayer (1975) investigated question type and its effect on achievement (categorized by question type outcomes). He found that certain types of questions may inhibit achievement as measured by other types (e.g., inserted calculating problems may inhibit achievement as measured by application problems). Threadgill (1979) found that inserted questions may benefit low ability students more than high ability students in studying mathematics text.

Most research on adjunct questions in text used immediate testing to measure learning with very few using delayed testing. Although test scores deteriorate over time, the facilitation of learning by the use of adjunct questions appears to hold (Dayton, 1977). Natkin and Stahler (1969) suggested that researchers should always include a delayed performance measure in their experiments.

While most adjunct question research has dealt with written material, there have been some investigations of their applications during lecture or through some alternative medium (e.g., film). McKenzie (1979) found that questions used as test-like events (i.e., addressed to a group) facilitated learning more so than questions addressed to individual students. Levine (1953) found that post-questions inserted during a film presentation are generally facilitating for students with low motivation. Sime and Boyce (1969) found that adjunct questions are superior to adjunct statements interspersed into a tape recorded lecture. The results of the use of test-like events during lecture or other media are not unanimous in favoring their use. Michael and Maccoby (1961) found no difference in general facilitating effects of test-like events used during a film presentation. Just as with adjunct question in text research, the focus of future research of test-like events in the classroom should be on the conditions of facilitating effects.

Purpose of the Study

The present study investigates the effectiveness of test-like events during lecture in the facilitation of learning mathematics. The study is based on the assumption that the use of test-like events (typically a problem to solve) during lecture, together with the provision of feedback, may increase the students' level of attention. This provision of feedback may aid students who give incorrect responses to the problems or those who give correct responses but are unsure about their correctness (Levie and Dickie, 1979).

Many instructional methods have been studied (e.g., individualized instruction, various discovery approaches, small group instruction, and electronic response systems), but either no significant advantage was found or, for reasons of practicality, such instructional methods failed to be implemented to any great extent (see Begle, 1979, for discussion). It seems unlikely that the lecture method will be supplanted by alternative approaches to any great extent, particularly at the college level. But perhaps by carefully studying what has been found in the literature on the use of test-like events, educational researchers can develop instruction that can be incorporated within the lecture (a condition of practicality). Most of past adjunct question research has been in highly controlled, laboratory settings with very short treatments. There is a need to move one step further along the research continuum, trying out certain conditions in a "normal" classroom, as suggested by Hilgard's (1964) continuum of pure research to technological development. Thus, the significance of the present study is its addition to the knowledge of the conditions for which test-like events will facilitate learning.

Theoretical Framework

The theory that the periodical use of test-like events during instruction will facilitate learning is based on the notion that such events will affect the level of student attention. The instructor controls nominal stimuli which can be arranged for potential reception by the student. It is the cognitive processing of the student that determines whether or not these potential nominal stimuli are received (Rothkopf, 1968). The procedure of students effectively processing the nominal stimuli is called attention (Anderson, 1970). Since the encoding of stimuli by the student is identified by inference rather than by direct observation, attention is a hypothetical construct that can only be indirectly manipulated and measured (Reynolds and Anderson, 1982). There are several aspects of attention to consider: length, intensity, selectivity, and nonselectivity.

The first of these, length of attention, is assumed to reflect the amount of time a student spends on an instructional task. Several studies (e.g., Frase, Patrick and Schumer, 1970; Watts and Anderson, 1971) found that test-like events increase the amount of time students spend on instructional materials.

Reynolds and Anderson (1982) found that test-like events in text increase the intensity of attention. The subjects in their study were told to press a key as quickly as possible whenever a tone sounded. It is postulated that when the subject is attending to some primary task, there will be a slight delay in responding to a secondary task. This implies that the cognitive processing capacity of the subject is limited, thus increasing the response time when attention is focused elsewhere. So to measure intensity of attention, Reynolds and Anderson used response time to a secondary task.

A third aspect of attention, selectivity, is of paramount importance to the present study. The selectivity of attention allows a person to follow one task or event while other tasks or events may be present simultaneously (Lindsay and Norman, 1977). Thus a student can choose to focus on a learning task while obscuring other things to which he or she might possibly attend. Attention selectivity can be jointly considered with either length or intensity of attention. Reynolds, Standiford, and Anderson (1979) found that test-like events in text increase the length of attention when subjects read text segments that contained information of the type addressed by the test-like events. Reynolds and Anderson (1982) reported that test-like events in text increased both length and intensity of attention when subjects were processing question relevant information. Thus, subjects selectively allocated more attention to material similar to the test-like events.

A fourth aspect of attention is nonselective, one for which a nonspecific heightening of vigilance occurs as opposed to focusing attention on specific types of information (Reynolds and Anderson, 1982). This is the aspect of attention that applies to Rothkopf's "indirect effects" of test-like events discussed earlier.

Frase (1970) suggested that test-like events produce both backward and forward processing effects for the learner, with each effect being either a selective or nonselective aspect of attention. A selective backward process leads the learner to mentally review material specifically related to the test-like event. A nonselective backward process leads the learner to review material adjacent to, but not necessarily directly related to the test-like event. A selective forward process leads the learner to focus on material similar to or specifically related to the test-like event during future instruction. A nonselective forward process leads the learner to attend to all material following the test-like event with greater intensity.

Test-like events produce various effects on the attention processes not only depending on several variables external to the learner, but also depending on the unique characteristics of the individual learner. Bovy (1981) proposed a cognitive information processing approach to instruction and discussed why instructional support, such as test-like events, may or may not facilitate learning. She discussed at great length how the effectiveness of instructional support depends on the processing skills of the individual learner. If the learner has a necessary processing skill and knows how and when to apply it at the appropriate stage of a learning task, then no external instructional support is required. However, if the learner has the necessary processing skill but lacks the aptitude to apply it at the appropriate stage of the learning task, then instructional support may activate the skill. The cognitive processing is actually conducted internally by the learner, but is directed by the instruction. Test-like events are a type of instructional support that may activate the processing of the learner, leading him or her to attend to relevant features of the learning task by requiring him or her to process specific elements of the material.

If the learner does not have the processing skill necessary for the learning task, a different type of instructional support may be necessary; one that externally provides the skill for the learner. For example, the use of manipulative aids in the learning of mathematics allows the learner to physically manipulate objects to compensate for the lack of visualization skills. Hence, the necessary processing skills are built into the instruction (commonly referred to as supplantation). Test-like events alone will not provide enough instructional support in this instance.

As the learning process proceeds over time, the nature of the attentive processes changes. Thus, the need for instructional support will change over time. As the learner becomes more sophisticated with regard to processing skills and the aptitude to apply them, less instructional support is needed. Another consideration of instructional support used for an extended period of time is the possibility of a long-term debilitating effect. By instruction continually performing the required cognitive operation, such as drawing attention to relevant aspects of a learning task, the learners may not develop that processing method for themselves (Bovy, 1981). What is needed is instructional support that can eventually be replaced by methods that require the learner to perform the cognitive task.

In summary, test-like events may facilitate learning by affecting some combination of length, intensity, selectivity, and nonselectivity of attention, while producing both backward and forward processing. The learner's cognitive processing skills and the aptitude to appropriately apply those skills determine the effectiveness of and the need for instructional support such as test-like events. Additionally, the long-term effects of such instructional support and the change in the learner's processing skills and aptitude over time must be considered.

Research Questions

The present study addresses the following research questions:

(1) Do skill level test-like events during mathematics lecture facilitate skill level learning?

(2) Do skill level test-like events plus concept/principle level test-like events facilitate skill level learning?

(3) Do skill level test-like events facilitate concept/principle learning?

(4) Do skill level test-like events plus concept/principle test-like events facilitate concept/principle learning?

(5) Does there exist a prior knowledge - treatment interaction?

(6) Does the use of test-like events produce any clear pattern of long-term effects or any change in effects over time?

Clearly there are many critical variables that may affect the effect of test-like events. The above research questions focus on the content of mathematics in the context of lecture. These and other critical variables (e.g., question type, student ability) will be addressed in the following review of the literature.

CHAPTER II

REVIEW OF THE LITERATURE

General Classroom Questioning

The use of questions during instruction is based on the same rationale as the use of test-like events. Typically, teachers will call on individual students to answer a question or will offer a rhetorical question in hope that the students' cognitive processes will be activated and focused on task. For questioning techniques to be effective, several conditions must be met. First, the student must attend to the question. Second, the student must interpret the meaning of the question. Often, the student will be confused by the meaning of the question because of the way the teacher phrases it. Since many teachers phrase questions spontaneously, some questions are poorly phrased (Gall, 1984). Finally, the student must generate a covert and sometimes an overt response. In order to do this, the student must have relevant information in memory and be able to process that information appropriately.

The theoretical framework of classroom questioning is similar to the aforementioned basis of test-like events. Classroom questioning may induce a practice and feedback effect. The questions give students an opportunity to practice recalling previous content and reflect upon it (backward processing). There may also be a cueing effect.

Questioning provides cues that may focus students' attention to relevant material or information (i.e., selectivity aspect of attention). Also, students develop an expectation of what is important for them to learn (i.e., an induced forward processing effect). And finally, there may possibly be a modality effect. The individual characteristics of the learner may determine the effectiveness of questioning, since it involves speaking and listening while seatwork involves reading and writing. For some students there may be less cognitive demand for speaking and listening (Gall, 1984).

Questioning during instruction has been a popular technique since the beginning of formal education; so it is quite natural for educational researchers to focus their attention on questioning to investigate its effectiveness in facilitating learning. Questioning research in the 19th and early 20th centuries focused on types of teacher questioning behavior and types of learning outcomes. There was a concern that only memory level thinking was being emphasized in the classroom (Wilen, 1984).

The interest in questioning research continued into the 1950s with more systematic and sophisticated research designs and instruments developed to analyze teacher questioning behavior and subsequent learning outcomes. The popularity of questioning research activity continued into the 1970s with the emphasis on identifying specific questioning levels and skills that affect student behavior and growth (Wilen, 1984). Most of this research focused on question type, usually fact versus higher cognitive. The common research question was: Do students learn more when teachers use factual questions or when they use higher cognitive questions (Gall, 1984)? Martin (1979) conducted a study investigating the effects of higher-order versus lower-order questions during class instruction on student processes and achievement. While higher-order questions had a significant effect on student processes (increased level of attention), they had no effect on achievement. Martin stated that this was consistent with most process-product studies. However, he did not describe any attempt to categorize question type on the criterion test. From the literature on adjunct questions in text one might hypothesize this to be an important factor (Mayer, 1975).

In a study of secondary level mathematics teachers, Evertson, Emmer, and Brophy (1980) reported that effective teachers (as measured by achievement) ask more product questions (short answer) and also more process questions (calling for an explanation) than do less effective teachers. More effective teachers also ask more new questions following correct responses.

Friedman (1976) categorized the question types used by geometry teachers into memory, comprehension, application, and higher level. He found that the majority of questions asked were comprehension level. This contrasts with other studies (Gall, 1970) which found that the majority of teachers' questions require simple recall. Thus, the content may, in part, determine what question types teachers emphasize.

One conclusion that can be drawn from past questioning research is that it is not a simple matter of concluding that factual questions increase learning or that higher cognitive questions increase learning. The interaction of question types and many other variables must be considered. Gall (1984) concluded that factual questions seem to be more effective for young low ability children in increasing achievement where the focus of instruction is on mastery of basic skills. Emphasis on higher cognitive questions seems more effective for average and high ability students, particularly as they reach high school age, where more independent thinking is required than in the primary grades.

Good and Brophy (1978) offered an explanation as to why factual questions may increase learning. They suggested that a teacher who frequently uses low level questions during instruction tends to be well organized. Thus, students are more focused on academic activities with fewer classroom management problems to distract from instructional tasks.

Whereas most questioning research has focused on question type and corresponding learning outcomes, several studies have dealt with the mechanics of teacher questioning. Morsh (Anderson, 1970) found that it is better for a teacher to ask a question, then call on a student to answer it, rather than calling on a student before asking the question. It is hypothesized that students will attend to the question better in the anticipation that they may be called upon to answer it.

Ryan (1974) found that both low level and high level achievement is increased when the number of students responding to teacher questioning is increased. Rather than calling on only one student for each question asked, the instructors of the low level questions groups and high level questions groups called on several students. With more students being actively involved in the instructional process, they may be more likely to attend to the questions posed. Ryan concluded that achievement is not solely a function of question type, but also a function of the questioning procedure.

Wait-time appears to be a crucial factor in enhancing the responses students give to questions. By extending the amount of time students are given to respond to questions, they respond more often and with greater elaboration (Rowe, 1974). Good (1981) found that junior high teachers have shorter wait-times for low achieving students.

Clute (1984) studied the interaction effect on achievement of mathematical anxiety and two instructional strategies. One strategy was a traditional, expository lecture. The other strategy was one of guiding students during instruction with questioning sequences addressed to individual students. The subjects of the study were students in a survey course in college mathematics. She found no instructional effect as measured by low level items (knowledge, comprehension) but a strong interaction effect with the expository method benefiting the high anxiety students and the questioning method benefiting those with low anxiety. However, Clute also found that when achievement was measured by high level items (application, analysis, synthesis, evaluation), the questioning method was superior across all anxiety levels. Also, no significant interaction was found when testing with high level items.

A few studies have examined questioning used as test-like events in the classroom. McKenzie and Henry (1979) compared the effects of using questions as test-like events directed to the entire class with questioning individual students during class instruction. The instruction consisted of a twenty minute class presentation. The researchers were interested in the effects on on-task behavior, test anxiety, and achievement. They found significant differences favoring the test-like events group in on-task behavior and achievement with no difference between the groups in test anxiety. The subjects were third-grade students and the response system used for the test-like events consisted of raising their hands versus not raising their hands (i.e., questions requiring dichotomous responses). The generalization of this study may not be all that broad considering that the treatment lasted only twenty minutes. McKenzie (1979) reported a similar study with similar results using different subject matter during the treatment.

Student Response Systems

During the 1960s and early 1970s, several articles were published describing the development and use of electronic student response systems. These devices allowed students to respond to an instructor's questions in the form of categorical data. Muller (1966) stated that the systems are based on the idea that feedback from the students during an instructional sequence allows the instructor to adjust the presentation of the material accordingly. For example, if a large percentage of the students give an incorrect response to an instructor's question, the instructor will go back and review the misunderstood material. Several of these articles simply describe the use and implementation of these electronic devices without reporting any controlled studies to validate their effectiveness (e.g., Monter, 1970; Littauer, 1972; Beach, 1974). However, several studies with varying degrees of control were reported. One study (Brown, 1972) investigated the effect of a student response system on achievement in a freshman level mathematics course. No difference was found in comparison to the traditional lecture method. It is not clear in the description of the study if the system was used merely to give the instructor feedback to determine if the students were following the presentation, or if test-like events were interspersed throughout the lecture. Also, Brown did not categorize the questions on the criterion test.

Chu (1972) did a thorough investigation of a student response system used at a college over a period of one and one-half years. During that time, 29 out of 140 faculty members chose to use the system. Generally, the instructors evaluated the devices very highly -- in fact, higher than did the students. The most common complaint or limitation reported by the instructors was the fact that the responses must be limited to categorical format (e.g., multiple choice questions). Chu reported one experiment, comparing the student response treatment to the traditional lecture method in the teaching of an undergraduate psychology course. The experimental group scored higher than the control group, although it was not statistically significant. The criterion test consisted mainly of factual questions. Bradley (Chu, 1972) reported that the use of the student response system increased learning as measured by quizzes but not the final examination, in the teaching of an education course. Rubin (1970) found no differences in quizzes, hour exams, or the final exam in comparing the student response treatment to the traditional lecture method in the teaching of agricultural economics classes. Casanova (1971) found the control group outperformed the student response group on problem sets, quizzes, hour exams and the final exam in organic chemistry. However, he reported that a disproportionate number of students in the control group withdrew from the course, leaving the results of the experiment inconclusive. Also, Casanova stated that 99% of the experimental group said they would prefer a response system class to a traditional one. Whitehead and Bassett (1975) found a five percent gain in achievement for a response system group in a human communications course. Bessler and Nisbet (1971) found no significant differences in achievement for a general education biology course (response system versus traditional lecture). The researchers used an ANCOVA design with the students' SAT composite score as the covariate.

Much of the research noted above found no gain in achievement when a student response system was utilized, while a few reported marginal gains. In all probability, there are conditions for which these electronic devices will increase learning. However, the focus of many of these studies seems to be on altering the behavior of the instructor to accomodate for student misunderstanding, rather than on increasing the attention level of the student. Also, these studies do not, in general, address question type, usually noting the limitations of a multiple choice format. Survey results noted in the studies show a generally favorable reaction to these systems by both instructors and students. However, one must ask if this is due mainly to novelty, with the favorable attitude fading after initial exposure. Finally, the cost of the systems and additional instructor preparation may be limiting factors, making it unlikely that the systems will be incorporated into the traditional lecture to any large degree.

Adjunct Questions in Text

While most adjunct question research has been in the context of learning from written material, not in the normal classroom setting, similar processes (i.e., attentional aspects) are hypothesized to result from an adjunct question treatment. The following is a discussion of what researchers have found with respect to the critical variables in research on adjunct questions in text. Particular attention is paid to the variables of question type and ability level and to studies utilizing mathematics subject matter. Other variables discussed which are relevant to the present study include motivation, knowledge of correct response and question placement.

In her review of adjunct questions in text research, Dayton (1977) stated that question type research has been fairly consistent in concluding that higher order questions tend to have a facilitating effect in later application of the concepts to problem solving situations. However, it may depend on what other interacting variables are being studied. Mayer (1975) studied question type effects in the context of mathematics text. He found that if only calculating adjunct questions were given, the subjects tended to perform well on the calculating problems on the criterion test but performed poorly on the conceptual questions (as compared to other treatment groups). He also found that those who were given only conceptual adjunct questions performed well on the conceptual questions on the criterion test but not as well on the calculating problems as compared to the calculating group. Subjects given all types of adjunct questions showed an overall superiority on the criterion test. Hence, it is possible that certain types of adjunct questions used exclusively can inhibit learning, depending on the type of learning measured. In a similar study, Mayer (1978) reported that when subjects received both calculating and comparative questions on a posttest, they performed relatively better on the type of questions they had received as adjunct questions. He suggested that comparative judgements of relations may require a different cognitive process than do calculations. He also suggested that past experience with solving only calculating problems in a low meaning context results in a "rigid encoding strategy," leading to poor transfer.

One of the problems faced in question type research is how one classifies the various questions. Wunderlich and Carry (1974) suggested more work is needed in developing appropriate levels of adjunct questions to fit the content being studied. This suggestion was based on their question type study which found knowledge only questions more beneficial than an all question types (knowledge, comprehension, and application) treatment, with no other differences found. The subject matter was mathematical functions with the treatment lasting 55 minutes. The subjects used were high school geometry students.

Other researchers have offered another perspective on question type research. After conducting seven experiments, Andre, Mueller, Womack, Smid and Tuttle (1980) concluded that there is no difference between application and factual adjunct questions in facilitating the application of concepts. They suggested that question type does not exert the powerful influence that educators have long believed. Their studies utilized psychology instructional materials and may not generalize to other content or settings under different conditions. For example, Watts and Anderson (1971) found that subjects who received adjunct questions requiring them to apply principles to examples not given in the main text performed better on a posttest composed of application questions than did subjects given adjunct questions requiring them to apply the same principles to examples they had seen in the main text. Watts and Anderson concluded that application adjunct questions induce students to process the text more thoroughly. However, the design of the experiment did not allow the researchers to attribute the effect to forward or backward processing.

As mentioned in Chapter I, the individual characteristics of the learner, in part, determine the effectiveness of instructional support such as adjunct questions. Some adjunct question studies have subjects stratified on the basis of measures of ability, aptitude or prior knowledge. Kuehls (1976) conducted a study investigating the effects of adjunct questions in mathematics text with first year calculus students serving as subjects. The content was matrix algebra and was chosen because of the lack of prior knowledge of the material on the part of the students. The questions were given in multiple choice format and no knowledge of results was given. The subjects were divided into levels (honors and nonhonors) and randomly assigned to treatments (adjunct questions and no questions). Kuehls found no statistically significant difference between the two treatment groups but did report a level-treatment interaction, with the nonhonors students benefiting more from the adjunct question treatment. Another study (Threadgill,

1979) examined the interaction between general aptitude, and two question types (applied versus verbatim) interspersed throughout mathematical text. The subjects were high school sophomores with the subject matter being logical implications. Application questions were found more effective, particularly with students at the lower end of the aptitude scale.

Shavelson, Berliner, Ravitch and Loeding (1974) reported similar findings with prose material. Subjects with low vocabulary knowledge benefited from higher order questions with high vocabulary subjects not benefiting from either question type. These results held for both an immediate recall test and a retention test given two weeks after the treatment.

The results of the above studies on ability level - adjunct question interaction are consistent with Dayton's (1977) review. She stated that while the literature is far from conclusive, it appears that low ability students tend to benefit more from adjunct questions than do high ability students. Perhaps high ability students already have an efficient learning strategy of their own and do not need instructional support such as adjunct questions.

The motivation level of the student appears to be another important variable in determining the effectiveness of adjunct questions. By manipulating monetary rewards, Frase, Patrick and Schumer (1970) found that adjunct questions were more beneficial for subjects with low motivation. The greater the reward for correct responses on the criterion test, the less effect the adjunct questions had on learning.

In another study utilizing mathematics text, Wiseman (1982) investigated the effects of question placement using the metric system as subject matter. The subjects were inservice elementary school teachers. No difference in intentional learning was found between the effects of adjunct questions given prior to instruction and those given after instruction. However, subjects given questions after instruction performed better on measures of incidental learning. These results are consistent with Rothkopf's findings that questions given prior to instruction may facilitate intentional learning but may inhibit incidental learning, whereas questions given after instruction may facilitate both types of learning. Rothkopf also found that knowledge of correct response may induce a forward processing effect leading the student to focus on the anticipated answer, disregarding other information (Anderson, 1970).

Little attention has been given to the comparison of adjunct questions in text to questioning techniques in a normal classroom setting. Rothkopf and Bloom (1970) addressed this comparison in a study where adjunct questions presented orally by a teacher were found to increase learning more so than did adjunct questions imbedded in text. They suggested that social interaction in the classroom can help shape and maintain effective study activities.

It is clear from the voluminous amount of adjunct question in text research that many other variables play a key role in determining the effectiveness of this type of instructional support. The age of the subjects may be one other important variable. The vast majority of adjunct question research studies have used high school or college students as subjects. This contrasts greatly with general classroom questioning research where the majority of studies have used younger students as subjects.

One other variable to consider is the medium in which instruction is given. This is addressed in the following section of this literature review.

Adjunct Questions in Other Media

The most thorough research on adjunct questions in media other than in written material is research on questioning effects during film presentation. Just as with questions in text, their effectiveness in film is dependent on many variables. Levie and Dickie (1973) reviewed the role of active response and feedback in media studies and concluded that overt responses are facilitating when the prescribed responses are relevant to the objective. The overtness of the response seems to be important. Vuke (1962) compared the effectiveness of an inserted question film to a film with no questions. The question group was given five seconds after each question to make a covert response. No significant differences were found between the groups or within four IQ subgroupings. The subjects were seventh grade general science students.

In a study of Air Force trainees viewing a film on world maps, Levine (1953) found that adjunct questions are facilitating only for learners with low motivation. This is consistent with the motivation findings of questions in text research.

In another alternative media study, Sime and Boyce (1969) found that the use of interspersed questions was more effective than interspersed statements during a thirty minute tape recorded lecture. The questions were presented on transparencies to undergraduate psychology students. The experiment was designed so that the difference found could be attributed to a generalized heightening of attention rather than feedback.

Although a different medium was used, the results of a study conducted by Teacher and Marchant (1974) conflict those of the Sime and Boyce study. Teacher and Marchant found that adjunct questions were not superior to adjunct statements with both treatments superior to a control. The questions and statements were interspersed in a film presentation.

Implications for the Present Study

From the review of the literature, there appear to be several sound reasons for believing that test-like events used during mathematics lecture might facilitate learning. In the situation where students are normally taking notes during lecture, it is a natural and convenient setting for students to make overt written responses to test-like events. Based on previous research, the opportunity to make overt responses rather than simply covert ones may be important (Levie and Dickie, 1973). Also, Ryan's (1974) study showing multiple student responses superior to a single student response lends support to the idea of test-like events during lecture.

With some teachers lacking questioning skills, as suggested by Gall (1984), test-like events in the classroom might offer a systematic, structured format allowing teachers to question students in a less spontaneous manner. The structure of the treatment seems to be one natural solution to the teacher wait-time problem. The instructor must allow students time to complete their overt responses to the test-like events, giving them more time to reflect upon the question and their answers.

The RothKopf and Bloom (1970) study of adjunct questions imbedded in text versus oral presentation suggests that test-like events during lecture may offer a social interaction that leads to effective learning in a way that a regular homework assignment or individual seatwork cannot. There may also be a modality effect of practicing during group instruction (i.e., interaction between the teacher and the students in a listening and speaking format rather than exclusively reading and writing when working individually).

The student response systems studies closely resemble a test-like events treatment during mathematics lecture, particularly the Brown (1972) study. However, this study, as with most response system studies, focused on the effect of the response system without regard to several important variables (e.g., question type and learner characteristics).

Based on the research on general classroom questioning, student response systems, and adjunct questions in text, together with the theoretical framework presented in Chapter I, the anticipated findings of the present study are discussed.

Research Question (1): Do skill level test-like events during mathematics lecture facilitate skill level learning? Mayer's (1975, 1978) studies of adjunct questions in mathematics text suggest that skill level questions will increase skill level learning. Students tend to focus on the question type. This could be a result of either a specific backward or a specific forward processing effect. From the research on general classroom questioning, these processing effects are not as well documented and the effect may not be quite as powerful as with questions imbedded in text. Thus, it is expected that skill level learning will be facilitated, but perhaps not to the degree found in questions in text research.

Research Question (2): Do skill level test-like events plus concept/principle test-like events facilitate skill level learning? Based on Mayer's (1975) study and other adjunct question research not focused on mathematics subject matter, one would expect the skill level learning of students receiving both question types to be greater than that of students receiving no test-like events but less than that of students receiving only skill level test-like events. With students receiving both types of problems, the selectivity factor of attention should not be as strong.

Research Question (3): Do skill level test-like events facilitate concept/principle learning? Based on adjunct question in text research, it is expected that skill level questions used exclusively will inhibit concept/principle learning rather than facilitate it. However, from general classroom questioning research, one might expect no significant effect either way. The degree of the selectivity role of attention in the treatment is the key factor in addressing this question, making it difficult to predict. Additionally, skills in mathematics are based on concepts and principles, making it quite plausible to expect the learning of one to facilitate the learning of the other. Hence, based on previous research and the relationship between skills, concepts and principles, it is not clear what to anticipate with respect to this research question.

Research Question (4): Do skill level test-like events plus concept/principle test-like events facilitate concept/principle learning? Similar to Research Question (2), it is expected that students receiving both types of questions will benefit more than students receiving no questions. It is also expected that students receiving both types of questions will score higher on concept/ principle measures than those students receiving only skill level questions.

Research Question (5): Does there exist a prior knowledge - treatment interaction? Several studies (e.g., Threadgill, 1979) investigated an aptitude or ability - treatment interaction, with fairly consistent results. Instructional support tends to be more effective for low ability (or aptitude) students. Whereas prior knowledge is partially a function of ability, it is expected that an interaction will exist with low prior knowledge students benefiting the most. Most studies have found the treatment to have no effect on high ability students; so it is expected that the interaction will be ordinal.

Research Question (6): With the use of four criterion tests interspersed throughout the treatment, does there exist any clear pattern of long-term effects or any change in effects over time? With the treatment of the present study lasting fifteen weeks, one must look at similar studies with long treatments to predict the effects of the treatment over time. The vast majority of the questions in text and in film research had very short treatments. However, many of the response systems studies had long treatments. The Brown (1972) study lasted an entire semester, but only reported results as measured by the final examination. No statistical difference was found. Bradley (Chu, 1972) reported that the use of a response system was more effective as measured by quizzes but not the final examination. Casanova (1971) found no differences on problem sets, quizzes, hour exams, or the final exam, but these results are inconclusive due to subject mortality. The McKenzie and Henry (1979) study of test-like events in the classroom closely parallels the present study except the treatment lasted only twenty minutes.

Integrating the results of past research, one might expect a test-like events treatment to have a short term effect, but to lose its effectiveness over time. This is consistent with the aspect of the theoretical framework relating to the change in learner characteristics over time, with the learner developing his or her aptitude at applying the appropriate processing skills. Hence, less instructional support is needed. On the other hand, a forward processing effect may be induced, raising the degree of the selectivity aspect of attention over time. If this is the case, it is expected that the effect of the treatment might increase over time for learning outcomes that parallel the question types in the treatment. Also, the nonselective aspect of attention must be considered. If the treatment is powerful enough to significantly affect this, it might be expected that the effect of the treatment would stay fairly consistent over the course of fifteen weeks.

Finally, several intervening variables may affect the answers to any or all of the six research questions. Important variables to consider are homework, knowledge from prior tests, expectations from previous mathematics courses, reading the textbook, or any activity outside of class on which students focus the task of learning the content of the course.

CHAPTER III

METHOD

Subjects

The unit of analysis in the present study is the university Calculus I student. The sixty-three subjects were students enrolled in Calculus I at the University of Oklahoma during the fall semester of 1984. The declared majors of the students serving as subjects were: twenty-one engineering, sixteen computer science, four engineering physics, three pre-medicine, three business, two geophysics, two meteorology, and eight other or no major. Forty-six of the subjects graduated from high schools in Oklahoma, with sixteen from out-of-state and one not reporting. Fifty-two of the subjects reported ACT scores, averaging 25.0 on the mathematics portion. Eleven subjects did not report ACT scores. There were forty-eight freshmen, ten sophomores, three juniors, one senior, and one graduate student.

Three sections of Calculus I, averaging forty students each, served as the initial pool of subjects. This pool of subjects was pretested for prior knowledge on the fourth day of class, before the treatment had been implemented. Only those students taking all four criterion tests were considered for final inclusion in the study.

The process of selecting subjects for the present study is an adaptation of the Cook and Campbell (1979) procedure of equating intact groups when random assignment is not possible. The partitioning of each section into a high prior knowledge cell and low prior knowledge cell and final selection of subjects was made to meet the following conditions:

(1) To equate subjects, based on prior knowledge scores, across high prior knowledge and low prior knowledge cells.

(2) To keep cell sizes proportional (i.e., avoid a nonorthogonal design), given condition (1).

(3) To maximize the final number of subjects analyzed in the study (i.e., maximize the power of the test statistics), given conditions (1) and (2).

(4) To keep cell sizes as equal as possible, given conditions (1) - (3).

(5) To equate cell error variances, given conditions (1) - (4).

To meet condition (1), students within each section with the highest or lowest prior knowledge scores were removed from consideration as subjects in the present study. Once condition (1) was met, subjects were randomly removed from each cell to meet conditions (2) - (4). Condition (5) proved intractable given conditions (1) - (4).

Sixty-three subjects were selected for analysis, with twenty-one in each section (i.e., treatment group). Within each section, eleven were ranked as high prior knowledge and ten were ranked as low prior knowledge. The dividing line between high and low prior knowledge blocks was the median score of the subjects selected (median = 39, 39 to 54 high, 19 to 38 low). Condition (1) was tested with one-way ANOVAs on the high prior knowledge groups and the low prior knowledge groups with $F(2,30) = 0.05$, $p = .956$ and $F(2,27) = 0.10$, $p = .903$

respectively (see Tables 1 - 3).

TABLE 1

PRETEST SAMPLE MEANS AND VARIANCES

Cell	Observations	Mean	Variance
High-control.	11	43.55	8.87
High-skill.	11	43.55	18.27
High-both	11	43.09	22.89
Low-control	10	30.80	38.07
Low-skill	10	29.90	37.43
Low-both.	10	31.00	29.78

TABLE 2

ANOVA, PRETEST MEANS, HIGH PRIOR KNOWLEDGE

Source of Variation	D.F.	Mean Square	F
Between Groups.	2	0.758	0.05
Within Error.	30	16.679	

TABLE 3

ANOVA, PRETEST MEANS, LOW PRIOR KNOWLEDGE

Source of Variation	D.F.	Mean Square	F
Between Groups.	2	3.433	0.10
Within Error.	27	33.426	

Treatment

General Description

The present study was conducted during the fall semester of 1984 at the University of Oklahoma. Three sections of Calculus I served as the treatment groups, with treatments being randomly assigned to sections. One section served as the control group and received no test-like events during lecture. A second section received only skill level test-like events. A third section received both skill and concept/principle test-like events. All three sections met for fifty minutes on Monday, Wednesday and Friday. The control, skill, and both question type groups met at 11:30 A.M., 1:30 P.M., and 2:30 P.M., respectively. The present researcher served as the instructor for all three sections.

For the two experimental groups, typically two or three test-like events were given during each lecture. The subjects in these groups were given time to make overt written responses. The amount of time allowed was at the discretion of the instructor, but typically was one to two minutes. The instructor then gave a correct response. The instruction given to the control group was the same as that for the experimental groups except that the problems used as test-like events were incorporated into the lecture.

All instruments used in the study were designed by the present researcher. The content validity of the pretest and all criterion tests was determined by a group of content specialists who judged them to be congruent with the domain of the subject matter. All test-like events, the criterion tests, and the pretest are given in Appendix A.

Variables of the Treatment

Question type. The test-like events used in the study were classified as either skill level or concept/principle level. Suydam and Dessart (1980) define skills as "... what students should be able to do. They are characterized in terms of proficiency, accuracy, or speed. When mastered, skills require relatively little reflection. They are based on concepts and principles." For the purpose of the present study, skill level problems included not only those problems fitting the above definition, but also those problems composed of one or more subordinate skills and commonly taught in a procedural or algorithmic fashion. This included problems such as max/min word problems, where the student is taught the procedure of finding a function in one variable to be maximized or minimized, finding the critical points of the function, etc. Other examples of skill level problems include:

(1) Given $f(x) = x/(2x+1)$, find $f'(3)$.

(2) Find an antiderivative of $5x + 8$.

(3) Evaluate $\lim_{x \to 2} (x-2)/(2-x)$.

For the purpose of the present study, concepts are used to mean defined and undefined mathematical terms. Examples of calculus concepts include: derivative, continuity, and instantaneous velocity. Evidence of concept learning includes (Sowder, 1980):

(1) Recognition of an example of a concept.

(2) Distinction between examples and nonexamples of a concept.

(3) Statement of a definition of a concept.

A principle is a relationship between two or more concepts.

Principles are used in this study to mean mathematical postulates and theorems. Examples of calculus principles include: the Mean Value Theorem, the Extreme Value Theorem, and the Intermediate Value Theorem. Evidence of principle learning includes:

(1) Classification of a situation as satisfying or not satisfying the "if" part of the principle.

(2) Generalization to the "then" part of the principle, given the "if" part.

(3) Recognition of the principle in a logically equivalent form (e.g., contrapositive form).

(4) Recognition of the converse of a principle and the ability to determine if it is itself a principle.

(5) Generation of specific examples of mathematical concepts that apply to a principle and ability to determine if no such examples exist.

(6) Statement of the principle.

Examples of concept/principle problems include:

(1) Why doesn't the Mean Value Theorem apply to $f(x) = x^{2/3}$, $x \in [-1,3]$?

(2) True/False. If the domain of a function is not a finite closed interval, then the range of the function is not a finite closed interval.

(3) What is implicit differentiation?

Each test-like event used during the treatment and each item on all criterion tests was categorized by the present researcher and two content specialists as either skill level or concept/principle level, based on the definitions and examples given above. The interrater reliabilities were computed (see Table 4). The classifications of all test-like events and all criterion test questions are listed in Appendix B, Tables 25 - 29.

TABLE 4

INTERRATER RELIABILITY OF CATEGORIZATION OF QUESTION TYPE

Treatment

Raters	Agree	Disagree	Percent Agree	Correlation
1 vs. 2	97	24	80.2%	.60
1 vs. 3	109	12	90.1%	.80
2 vs. 3	103	18	85.1%	.71

Exam I

Raters	Agree	Disagree	Percent Agree	Correlation
1 vs. 2	27	3	90.0%	.82
1 vs. 3	29	1	96.7%	.94
2 vs. 3	28	2	93.3%	.87

Exam II

Raters	Agree	Disagree	Percent Agree	Correlation
1 vs. 2	18	1	94.7%	.88
1 vs. 3	17	2	89.5%	.80
2 vs. 3	16	3	84.2%	.70

TABLE 4 - Continued

Exam III

Raters	Agree	Disagree	Percent Agree	Correlation
1 vs. 2	26	1	96.3%	.90
1 vs. 3	26	1	96.3%	.90
2 vs. 3	27	0	100.0%	1.00

Final Exam

Raters	Agree	Disagree	Percent Agree	Correlation
1 vs. 2	45	1	97.8%	.95
1 vs. 3	46	0	100.0%	1.00
2 vs. 3	45	1	97.8%	.95

There was not unanimous agreement among the raters as to how the test-like events and exam questions should be classified. This tends to be relative to the instruction and prior experience of the student. If instruction has provided the student with an algorithmic procedure to complete a task, then the task would be classified as a skill problem. If, on the other hand, the student does not have an algorithmic device, but has been provided with the concepts necessary to complete the task, then it would be classified as a concept/principle problem.

This problem of classification is similar to that of classifying a task as completing an exercise or solving a problem. That is, it is relative to the experience of the learner (Lester, 1980). Since the present researcher (Rater 3) was the instructor of the treatment groups, thus aware of what algorithmic procedures were taught, his classifications were used in the analysis of this study.

Materials and content. The textbook used was Calculus and Analytic Geometry, (Gillett, 1984). The material used from the textbook came from chapters two through six, plus the first two sections of chapter seven. Sections 2.4 (limits) and 5.4 (Newton's method) were omitted. Chapter two covered the concept of derivative, instantaneous velocity, slope at a point, a review of trigonometry, derivatives of trigonometric functions, and properties of limits. Chapter three covered the techniques of differentiation (power, product, quotient, chain rules) and implicit differentiation. Chapter four covered continuity, the Mean Value Theorem and its applications, higher order derivatives, linear approximation and differentials. Chapter five covered rate of change, extreme values, curve sketching and max/min problems. Chapter six covered area under a curve, the integral, the Fundamental Theorem of Calculus, and integration by substitution. The first two sections of chapter seven covered area between curves and volume of a solid of revolution.

Examination I tested the material from chapters two and three, with Examination II testing chapters four and five, and Examination III testing chapter six plus the first two sections of chapter seven. The Final Examination was comprehensive. The subjects were given fifty minutes to complete Exams I, II, and III, and two hours to complete the Final Exam.

Question placement. Previous research (e.g., Rothkopf, 1966) has shown that test-like events given prior to instruction may facilitate intentional learning but may inhibit incidental learning. Based on this research, the present researcher chose to design the treatment so that all test-like events would follow instruction (i.e., follow the development of the skill, concept or principle on which the test-like event was based).

Question frequency. Typically, two or three test-like events were given to the experimental groups during each fifty minute lecture. Question frequency for the study was determined by what previous research has found and also to keep conditions equivalent across treatment groups. According to Natkin and Stahler (1969), when long-term performance is of primary concern, questions should be used sparingly, perhaps only at quite important points. Otherwise, the test-like events will not induce high arousal (i.e., nonselective level of attention). Additionally, too many test-like events given during a lecture would consume a significant amount of time. Thus, material covered in the experimental groups would not be the same as that covered in the control group.

Knowledge of correct response. All test-like events given during the treatment were followed by correct responses from the instructor. The treatment was designed to include correct responses based on the findings of research on feedback. When feedback follows a correct response, it tells the learner that his or her strategy is achieving the desired goal of transferring information. In terms of achievement, supplying feedback after an incorrect response is probably more important than for a correct response. Feedback not only informs the student that the response is incorrect, but may also substitute correct information in place of incorrect information. However, there are some instances when feedback may have a negative effect on incidental learning. The test-like events may induce a forward processing effect leading the student to focus on the anticipated answer, thus, disregarding other information (Anderson, 1970).

Medium. All test-like events and corresponding correct responses were given by the instructor using an overhead projector and/or the chalkboard. The correct responses were usually elaborated on orally.

Data Analysis

A multivariate analysis of variance (MANOVA) is an appropriate test for an experiment involving several dependent variables. The MANOVA is used to investigate the effects on means of the experimental treatments on all dependent variables considered simultaneously (Hays, 1981). It is used to protect the Type I error rate for several univariate tests (ANOVAs) through use of a "step-down" procedure. The univariate tests are performed if and only if the MANOVA test is significant. The multivariate stepdown procedure and the Dunn-Bonferroni multiple univariate procedure both protect against an inflated Type I error rate. However, if the measures of the dependent variables are correlated, as is expected in the present study, the multivariate step-down procedure is more powerful.

It is possible to have a multivariate test that is significant, with no univariate test signficant. That is, the multivariate test can possibly detect significant differences attributed to effects on some linear combination of the dependent variables while not detecting any differences for any individual dependent variable. One consequence of this is the difficulty of interpreting the results in such a situation.

For the present study, the following multivariate stepdown procedure was used:

Step (1). A Hotelling-Lawley Trace 3X2 (treatment X prior knowledge) MANOVA test for the null hypotheses of no overall treatment effect and no overall treatment X prior knowledge effect is performed. Each hypothesis is tested at the .05 level of significance. If both null hypotheses are accepted then conclude that there are no significant treatment or interaction effects for any of the dependent variables. If one or both of the null hypotheses are rejected, proceed to step (2).

Step (2). All eight univariate ANOVAs are performed. If all null hypotheses of no treatment and no interaction effects (each tested at the .05 level of significance) are accepted, then conclude that there is an overall effect on some linear combination of the dependent variables, but not for any individual dependent variable. If any of the null hypotheses are rejected, proceed to step (3).

Step (3). Post hoc individual comparison procedures are performed on any ANOVAs resulting in significant treatment effects or interaction effects. Since the present study has three treatment groups, the Newman-Keuls multiple range test will control Type I error (.05) experimentwise for each dependent variable. Hence, it is appropriate for testing of treatment effects. An appropriate individual comparison test on interaction is performed on all ANOVAs resulting in significant interactions.

After the multivariate stepdown procedure was performed, the six research questions were addressed as follows (some may have been rendered moot by the stepdown procedure):

(1) Do skill level test-like events facilitate skill level learning? This question was analyzed for each of the four measures of skill learning by comparing the skill only group to each of the other two groups.

(2) Do skill level test-like events plus concept/principle test-like events facilitate skill level learning? This question was analyzed for each of the four measures of skill learning by comparing the skill plus concept/principle group to each of the other two groups.

(3) Do skill level test-like events facilitate concept/principle learning? This question was analyzed for each of the four measures of concept/principle learning by comparing the skill only group to each of the other two groups.

(4) Do skill level test-like events plus concept/principle test-like events facilitate concept/principle learning? This question was analyzed for each of the four measures of concept/principle learning by comparing the skill plus concept/principle group to each of the other two groups.

(5) Does there exist a prior knowledge - treatment interaction? This question was analyzed for all eight dependent variables by the MANOVA interaction test and any subsequent ANOVA and individual comparison interaction tests.

(6) Does the use of test-like events produce any clear pattern of long-term effects or any change in effects over time? This question was analyzed by observing (rather than using a statistical test) the change in effects across measures of both skill and concept/principle learning.

CHAPTER IV

RESULTS

The Stepdown Procedure

The analysis of the data using the MANOVA stepdown procedure was carried out as stated in Chapter III. The statistical package employed was SAS (SAS Institute Inc., 1982).

Step one. Using the Hotelling Lawley trace, the MANOVA null hypothesis of no overall treatment effect was rejected, $Tr(E^{-1}H) = 0.59$, $F(16,98) = 1.81$, $p = 0.04$. The MANOVA null hypothesis of no overall treatment X block interaction was not rejected, $Tr(E^{-1}H) = 0.15$, $F(16,98) = 0.47$, $p = 0.95$.

In accordance with the stepdown procedure, it was concluded that there are no treatment X prior knowledge interaction effects. Since the MANOVA test revealed an overall treatment effect, the eight univariate tests (ANOVAs) were performed and analyzed, to investigate possible treatment effects for each of the dependent variables. The eight ANOVAs are presented in Tables 5 through 12.

Step two. For concept/principle learning, Exam I, the test of the main effect for treatment was significant, $F(2,57) = 3.66$, $p = 0.03$.

For skill learning, Exam I, the test of the main effect for treatment was not significant, $F(2,57) = 2.09$, $p = 0.33$.

For concept/principle learning, Exam II, the test of the main effect for treatment was significant, $F(2,57) = 4.73$, $p = 0.01$.

For skill learning, Exam II, the test of the main effect for treatment was not significant, $F(2,57) = 1.09$, $p = 0.34$.

For concept/principle learning, Exam III, the test of the main effect for treatment was significant, $F(2,57) = 4.59$, $p = 0.01$.

For skill learning, Exam III, the test of the main effect for treatment was not significant, $F(2,57) = 0.58$, $p = 0.56$.

For concept/principle learning, Final Exam, the test of the main effect for treatment was not significant, $F(2,57) = 0.97$, $p = 0.39$.

For skill learning, Final Exam, the test of the main effect for treatment was not significant, $F(2,57) = 1.08$, $p = 0.35$.

The analyses of block effects are also reported in Tables 5 through 12, although no research question directly pertained to them. Additionally, the univariate interaction analyses are included although the conclusion of no significant interaction was made at step one.

Step three. Three of the eight univariate tests revealed significant main effects for treatment. In accordance with the stepdown procedure, the Newman-Keuls individual comparison procedure was performed on all pairwise comparisons (at ALPHA = 0.05 for each dependent variable) for the three dependent variables revealing significant mean differences. All comparisons performed are reported in Tables 13 through 15.

For concept/principle learning, Exam I, the group receiving both question types scored significantly higher than did the skill only group. The other two pairwise comparisons revealed no significant differences.

For concept/principle learning, Exam II, the both question type group scored significantly higher than both the control and the skill only groups. There was no significant difference between the control and skill groups.

For concept/principle learning, Exam III, the both question type group scored significantly higher than did the skill only group. No other comparisons were significant.

The observed means and standard deviations for all eight dependent variables are reported by cells and by treatment groups in Tables 16 through 23.

TABLE 5

ANOVA, CONCEPT/PRINCIPLE LEARNING, EXAM I

Source	D.F.	Sum of Squares	F
Treatment	2	26.41	3.66*
Block	1	11.07	3.06
Interaction	2	5.56	0.77
Error	57	205.94	

*Significant at .05 level.

TABLE 6

ANOVA, SKILL LEARNING, EXAM I

Source	D.F.	Sum of Squares	F
Treatment	2	129.75	2.09
Block	1	675.74	21.79***
Interaction	2	1.17	0.02
Error	57	1767.95	

***Significant at .001 level.

TABLE 7

ANOVA, CONCEPT/PRINCIPLE LEARNING, EXAM II

Source	D.F.	Sum of Squares	F
Treatment	2	24.03	4.73*
Block	1	2.71	1.07
Interaction	2	2.16	0.42
Error	57	144.75	

*Significant at .05 level.

TABLE 8

ANOVA, SKILL LEARNING, EXAM II

Source	D.F.	Sum of Squares	F
Treatment	2	84.95	1.09
Block	1	523.67	13.47***
Interaction	2	46.56	0.60
Error	57	2216.53	

***Significant at .001 level.

TABLE 9

ANOVA, CONCEPT/PRINCIPLE LEARNING, EXAM III

Source	D.F.	Sum of Squares	F
Treatment	2	58.51	4.59*
Block	1	55.29	8.68**
Interaction	2	8.63	0.68
Error	57	363.23	

*Significant at .05 level.
**Significant at .01 level.

TABLE 10

ANOVA, SKILL LEARNING, EXAM III

Source	D.F.	Sum of Squares	F
Treatment	2	25.52	0.58
Block	1	607.61	27.65***
Interaction	2	20.09	0.46
Error	57	1252.49	

***Significant at .001 level.

TABLE 11

ANOVA, CONCEPT/PRINCIPLE LEARNING, FINAL EXAM

Source	D.F.	Sum of Squares	F
Treatment	2	53.56	0.97
Block	1	322.51	11.84**
Interaction	2	22.49	0.41
Error	57	1579.08	

**Significant at .01 level.

TABLE 12

ANOVA, SKILL LEARNING, FINAL EXAM

Source	D.F.	Sum of Squares	F
Treatment	2	272.79	1.08
Block	1	2826.76	22.35***
Interaction	2	25.16	0.10
Error	57	7210.36	

***Significant at .001 level.

TABLE 13

NEWMAN-KEULS TEST FOR CONCEPT/PRINCIPLE LEARNING, EXAM I, TREATMENT EFFECTS

Alpha = 0.05 DF = 57 Mean Square Error = 3.61

Comparison	Means
Control - Skill	(10.62 - 10.24)
Control - Both.	(10.62 - 11.76)
Skill - Both.	(10.24 - 11.76)*

*Significantly different.

TABLE 14

NEWMAN-KEULS TEST FOR CONCEPT/PRINCIPLE LEARNING, EXAM II, TREATMENT EFFECTS

Alpha = 0.05	DF = 57	Mean Square Error = 2.54

Comparison	Means
Control - Skill	(5.86 - 5.90)
Control - Both.	(5.86 - 7.19)*
Skill - Both.	(5.90 - 7.19)*

*Significantly different.

TABLE 15

NEWMAN-KEULS TEST FOR CONCEPT/PRINCIPLE LEARNING, EXAM III, TREATMENT EFFECTS

Alpha = 0.05	DF = 57	Mean Square Error = 6.37

Comparison	Means
Control - Skill	(15.48 - 14.62)
Control - Both.	(15.48 - 16.95)
Skill - Both.	(14.62 - 16.95)*

*Significantly different.

TABLE 16

OBSERVED MEANS AND STANDARD DEVIATIONS
FOR CONCEPT/PRINCIPLE LEARNING, EXAM I

By Cell	Observations	Mean	S.D.
High-Control.	11	10.73	1.79
High-Skill.	11	10.54	1.44
High-Both	11	12.55	1.51
Low-Control	10	10.50	2.22
Low-Skill	10	9.90	1.97
Low-Both.	10	10.90	2.38
By Treatment	**Observations**	**Mean**	**S.D.**
Control	21	10.62	1.96
Skill	21	10.24	1.70
Both.	21	11.76	2.10

TABLE 17

OBSERVED MEANS AND STANDARD DEVIATIONS
FOR SKILL LEARNING, EXAM I

By Cell	Observations	Mean	S.D.
High-Control.	11	29.27	3.58
High-Skill.	11	30.27	3.72
High-Both	11	32.73	1.56
Low-Control	10	22.50	6.17
Low-Skill	10	24.10	7.40
Low-Both.	10	26.00	8.45
By Treatment	**Observations**	**Mean**	**S.D.**
Control	21	26.05	5.96
Skill	21	27.33	6.44
Both.	21	29.52	6.72

TABLE 18

OBSERVED MEANS AND STANDARD DEVIATIONS FOR CONCEPT/PRINCIPLE LEARNING, EXAM II

By Cell	Observations	Mean	S.D.
High-Control.	11	5.91	2.02
High-Skill.	11	6.00	1.61
High-Both	11	7.64	1.69
Low-Control	10	5.80	1.48
Low-Skill	10	5.80	1.55
Low-Both.	10	6.70	0.95
By Treatment	**Observations**	**Mean**	**S.D.**
Control	21	5.86	1.74
Skill	21	5.90	1.55
Both.	21	7.19	1.44

TABLE 19

OBSERVED MEANS AND STANDARD DEVIATIONS
FOR SKILL LEARNING, EXAM II

By Cell	Observations	Mean	S.D.
High-Control.	11	26.55	4.55
High-Skill.	11	27.73	3.64
High-Both	11	30.55	5.01
Low-Control	10	23.20	9.26
Low-Skill	10	20.90	7.22
Low-Both.	10	23.40	6.55
By Treatment	**Observations**	**Mean**	**S.D.**
Control	21	24.95	7.20
Skill	21	24.48	6.50
Both.	21	27.14	6.73

TABLE 20

OBSERVED MEANS AND STANDARD DEVIATIONS
FOR CONCEPT/PRINCIPLE LEARNING, EXAM III

By Cell	Observations	Mean	S.D.
High-Control.	11	15.91	1.87
High-Skill.	11	15.91	2.84
High-Both	11	17.91	1.45
Low-Control	10	15.00	3.62
Low-Skill	10	13.20	2.70
Low-Both.	10	15.90	2.18
By Treatment	**Observations**	**Mean**	**S.D.**
Control	21	15.48	2.80
Skill	21	14.62	3.04
Both.	21	16.95	2.06

TABLE 21

OBSERVED MEANS AND STANDARD DEVIATIONS
FOR SKILL LEARNING, EXAM III

By Cell	Observations	Mean	S.D.
High-Control.	11	27.18	1.54
High-Skill.	11	27.91	2.46
High-Both	11	28.36	2.46
Low-Control	10	21.90	5.00
Low-Skill	10	20.10	7.98
Low-Both.	10	22.80	6.11
By Treatment	**Observations**	**Mean**	**S.D.**
Control	21	24.67	4.44
Skill	21	24.19	6.81
Both.	21	25.71	5.28

TABLE 22

OBSERVED MEANS AND STANDARD DEVIATIONS
FOR CONCEPT/PRINCIPLE LEARNING, FINAL EXAM

By Cell	Observations	Mean	S.D.
High-Control.	11	29.18	2.60
High-Skill.	11	27.09	6.17
High-Both	11	29.82	2.27
Low-Control	10	25.60	6.74
Low-Skill	10	23.30	6.86
Low-Both.	10	23.60	5.23
By Treatment	**Observations**	**Mean**	**S.D.**
Control	21	27.48	5.21
Skill	21	25.29	6.63
Both.	21	26.86	5.00

TABLE 23

OBSERVED MEANS AND STANDARD DEVIATIONS
FOR SKILL LEARNING, FINAL EXAM

By Cell	Observations	Mean	S.D.
High-Control.	11	55.45	5.20
High-Skill.	11	54.18	7.03
High-Both	11	59.00	5.62
Low-Control	10	43.80	12.32
Low-Skill	10	39.60	19.21
Low-Both.	10	45.00	12.66
By Treatment	**Observations**	**Mean**	**S.D.**
Control	21	49.90	10.83
Skill	21	47.24	15.70
Both.	21	52.33	11.80

Scorer Reliability

One source of error variance in the present study is the variance in the scoring of the criterion tests. For most of the test items, partial credit was possible. Thus, one would expect that there would be some variation in the test scores had a different grader been used.

Scorer reliability is found by having two independent scorers grade a sample of tests. The two scores obtained for each subject are correlated, with the correlation coefficient being the measure of scorer reliability (Anastasi, 1968).

In the present study five subjects from each of the six cells were randomly selected for the purpose of calculating scorer reliability. The present researcher and a content specialist independent of the study graded all four criterion tests. The scorer reliability was then calculated for all eight dependent variables. The correlations, given in Table 24, ranged from .86 for concept/principle learning, Final Exam, to .99 for concept/principle learning, Exam III.

TABLE 24

SCORER RELIABILITY

Instrument	Correlation
Concept/Principle, Exam I	.92
Skill, Exam I	.95
Concept/Principle, Exam II.	.96
Skill, Exam II.	.96
Concept/Principle, Exam III	.99
Skill, Exam III	.96
Concept/Principle, Final Exam	.86
Skill, Final Exam	.92

Results with Respect to Research Questions

Research question (1): Do skill level test-like events facilitate skill level learning? For all four measures of skill level learning, the both question type group scored higher than the skill only group. However, the differences were not significant. The control group scored higher than the skill only group on three of the four skill measures but not significantly higher. For the one skill measure on which the skill group outscored the control group, the difference was not significant.

Research question (2): Do skill level test-like events plus concept/principle test-like events facilitate skill learning? The both question type group scored higher than the control and skill only groups on all four measures of skill learning but not significantly higher.

Research question (3): Do skill level test-like events facilitate concept/principle learning? The control group scored higher than the skill only group on three of the four measures of concept/principle learning, but not significantly higher. For the one measure of concept/principle learning on which the skill group outscored the control group, the difference was not significant. The both question type group outscored the skill group on all four measures of concept/principle learning with three of the four differences significant.

Research question (4): Do skill level test-like events plus concept/principle test-like events facilitate concept/principle learning? The both question type group scored higher than the control group on three of the four measures of concept/principle learning with only one of the three being significantly higher. For the one measure of concept/principle learning on which the control group outscored the both question type group, the difference was not significant. The both question type group outscored the skill group on all four concept/ principle measures, with three of the four differences significant.

Research question (5): Does there exist a prior knowledge - treatment interaction? The results show no significant (or even marginal) interactions for all eight dependent variables.

Research question (6): Does the use of test-like events produce any clear pattern of long-term effects or any change in effects over time? For skill level learning, the both question type group outscored the control and skill groups on all four measures with no differences significant. The control group outscored the skill group on skill learning, Exam II, Exam III and the Final Exam. No differences were significant.

For concept/principle learning, the both question type group outscored the skill group on all four measures. The differences were significant on Exams I, II, and III. The difference was not significant on the Final Exam. The both question group outscored the control group on Exams I, II, and III, with significance only on Exam II. The control group outscored the both question type group on the Final Exam, but not significantly. The control group outscored the skill group on Exam I, Exam III, and the Final Exam, though not significantly. The skill group outscored the control group on Exam II, but not significantly.

To summarize the trend of the significant results, none were found with respect to skill learning across all four skill measures. Significant differences were found with respect to concept/principle learning on Exams I, II, and III, with the both question group scoring significantly higher than the skill group, but significantly higher than the control group only on Exam II. No significant differences were found on the Final Exam. As mentioned with respect to research question (5), there were no significant interactions across all skill and concept/principle measures.

CHAPTER V

DISCUSSION

Summary of Results

Skill Learning

For the three treatment groups of the study (control, skill only test-like events, and skill plus concept/principle test-like events), no statistically significant differences were found on the four measures of skill level learning (from Exams I, II, III, and the Final Exam). The group given both question types outperformed (nonsignificant) the other two groups on all four measures of skill learning with the control group exceeding (nonsignificant) the skill only group on three of the four skill measures. Only on Exam I did the skill only group exceed (nonsignificant) the control group on skill learning.

Concept/Principle Learning

Four pairwise comparisons were found to be statistically significant on the four measures of concept/principle learning. The group receiving both question types outperformed the skill only group on the measures of concept/principle learning from Exams I, II, and III. The both question type group also outperformed the control group on concept/principle learning, Exam II.

For comparisons that were not statistically significant, the both question type group outperformed the skill only group on concept/principle learning, Final Exam. The both question type group outperformed the control group on concept/principle learning, Exams I and III, with the control group exceeding the both question type group on the Final Exam. The control group exceeded the skill only group on concept/principle learning, Exam I, Exam III, and the Final Exam. Only on Exam II did the skill group exceed the control group.

Treatment - Prior Knowledge Interaction

No statistically significant treatment - prior knowledge interaction effects were found for any of the eight measures of learning.

The Results as Related to Past Findings

The fact that no statistically significant advantage was found for either the skill only group or the skill plus concept/principle group with respect to all four measures of skill level learning is clearly incongruous with most studies focusing on question type in mathematics text. On the basis of Mayer's (1975, 1978) studies one would predict that the skill only group would outperform both the control group and the both question type group and that the both question type group would outperform the control group. One possible explanation for this difference is based on the theoretical framework as outlined by Bovy (1981). If the learner has the necessary processing skill and knows how and when to apply it, no external instructional support is needed. Calculus I students have had several years of mathematics and with the exception of a plane geometry course, the focus of the previous instruction has most likely been on skill development. Thus, a forward processing effect has been induced by previous mathematics instruction, leading the students to focus on skill level problems during instruction. Hence, Calculus I students already have the skill and aptitude to attend to skill level problems during instruction.

The statistically significant differences found with respect to concept/principle learning are in partial agreement with Mayer's findings. Based on his findings, one would expect the both question type group to outperform both the control and skill only groups and the control group to outperform the skill only group. The both question type group did outperform ($p < 0.05$) the control group on one of the four concept/principle measurements and outperformed ($p < 0.05$) the skill only group on three of the four concept/principle measurements. The fact that the both question group outperformed the control group on one concept/principle measurement offers some support to the notion that the subjects may not have the aptitude to focus on concept/principle problems during instruction, so that the concept/principle test-like events provided a facilitating instructional support. The fact that the both question type group outperformed the skill only group on three of the four measurements may be explained by Mayer's description of the selective attention process. He stated that test-like events may induce a selective attention process leading the learner to reduce the amount of irrelevant information attended to concerning the goals of instruction. Overemphasis on some types of test-like events may limit the acquisition of other important aspects of instruction. Thus, the subjects of the skill only group may have overfocused on skill level problems thereby limiting their focus on concept/principle problems. However, this explanation is only partially satisfactory since the control group did not significantly outperform the skill only group on concept/principle measurements. It is not entirely clear if the differences found lead to a conclusion of the both question type treatment facilitating concept/principle learning or the skill only treatment inhibiting concept/principle learning.

The findings of no prior knowledge - treatment interaction are not in agreement with past findings involving aptitude or ability - treatment interaction. Based on Threadgill's (1979) study of interaction effects using test-like events in mathematics text one would expect a significant interaction with a test-like events treatment benefiting the low prior knowledge subjects more than those with high prior knowledge. The subjects used in the Threadgill study were high school sophomores who were blocked according to a general aptitude measurement. There are at least two explanations for the difference in the findings of the present study and the Threadgill study. One possibility is the problem of finding an appropriate blocking variable, one that has often plagued aptitude treatment interaction research. Perhaps prior knowledge was not a suitable blocking variable for the present study. Another possibility is the fact that the variation within each group of subjects was not enough to detect an interaction effect. Since students from the extreme ends of the prior knowledge scale were removed from the data analysis in order to equate prior knowledge cell means, the students most likely to create an interaction effect may not have been analyzed. Also, the nature of the abilities and aptitudes of Calculus I students is probably different from the subjects used in Threadgill's investigation.

Clearly there are many other differences in the settings of the present study and previous studies of test-like events in mathematics text. One important difference is the fact that the present study was conducted in a normal classroom setting, whereas studies such as Mayer's were conducted in a laboratory setting. The subjects of Mayer's studies were psychology students participating in experiments to fulfill a course requirement, whereas the subjects of the present study were students enrolled in the calculus classes used in the treatment. The results of the criterion tests directly determined the subjects' grades in the present study but not in Mayer's studies. One plausible consequence of this difference is the factor of motivation. Frase et al. (1970) suggested that the advantage of test-like events diminishes as motivation increases. This is not to say that test-like events inhibit learning under conditions of high motivation but rather that the control group improves its performance relative to the experimental groups. Thus, if the subjects of the present study have high incentive relative to subjects participating in an experiment to fulfill a course requirement, one would expect a smaller effect for the present study. This is consistent with the film study of Levine (1953) which found that active review helps only students with low motivation. Anderson (1970), in a review of the literature on testlike events during film presentation, suggested that if the content is intrinsically interesting then the students will pay close attention even without test-like events.

Thus, one must consider the nature of the Calculus I student. With the majority of them majoring in engineering, computer science, or one of the physical sciences, it is quite plausible that they are students who find the subject matter interesting.

Studies which are more analogous to the present study, not only in terms of motivation but also in terms of length of treatment, are the student response system investigations. Brown (1972) found no significant difference in achievement as measured by the final examination when comparing the traditional lecture to the response system treatment, but found that subjective student evaluations suggested that the response systems reduced mathematics anxiety. His findings are consistent with the Final Exam results of the present study although his questions were not classified by type and no data analysis was undertaken on unit examinations.

Evaluations were also filled out by the students in the three sections of the present study, with the both question type section rating both the course and the instructor higher than did the skill only section, and the skill only section rating both the course and the instructor higher than did the control group. The scale used in these evaluations is a five point scale ranging from far above average to far below average. The both question type section rated the course 0.5 standard deviations higher than did the skill only section which rated the course 0.6 standard deviations higher than did the control section. The both question type section rated the instructor 0.4 standard deviations higher than did the skill only section which rated the instructor 0.5 standard deviations higher than did the control section. The interpretation of these student evaluations should be done with extreme caution since they include some students who were not used in the final data analysis. The student evaluation response distributions are given in Appendix B, Table 31.

In another response system study, Rubin (1970) found no differences between a traditional lecture treatment and a response system treatment as measured by quizzes, hour exams and the final exam. He did not classify questions by type either in the treatment or on the criterion tests. Rubin suggested that classroom activities, such as students responding to questions during instruction, are irrelevant to test performance since the student's ability to study and perform on examinations may outweigh all other variables. While the student's ability to study and perform on examinations may be the most important variable (one that would include prior knowledge and internal motivation), this argument in its extreme form would lead to a conclusion that all forms of instruction produce equivalent results.

Rubin's point does, however, lead to an explanation as to why no significant differences were found in the present study on concept/principle learning, Final Exam, but were found on Exams I, II, and III. The students in the control and skill only groups may have attended to the concept/principle questions on Exams I, II, and III when they studied for the Final Exam. That is, the unit examinations may have induced a forward processing effect resulting in these students compensating for the lack of concept/principle test-like events. This is consistent with the findings of Bradley's (Chu, 1972) response system study in which significant differences were found on unit examinations but not on a comprehensive final examination.

In Casanova's (1971) discussion of why a response system treatment may not be found superior to a traditional treatment, he suggested that overall performance of all treatment groups may be superior relative to other comparable classes because of the degree of preparation and coordination required in an experiment. He appears to be arguing that the amount of instructional preparation is the more critical variable, rather than the independent variable in the study. The present researcher feels that this explanation is a bit too convenient any time significant differences are not found.

The structure of the treatments for the experimental and control groups in the present study is very similar to that of the Sime and Boyce (1969) study comparing an adjunct question treatment to one with adjunct statements during a tape recorded lecture. The control group in the present study was given the same content as the experimental groups except the test-like events were incorporated into the lecture for the control group (as were the concept/principle questions in the skill only group). Thus, the test-like events were essentially presented to them as information containing both the question and the answer. Sime and Boyce concluded that while the superiority of the adjunct question treatment is not large, the question format provides "something extra," over and above the direct imparting of information contained in the adjunct statement. This was partially the case for concept/principle learning in the present study.

Some of the past questioning research tried to measure attention as a process variable to see if it coincides with achievement as a product variable. For example, in the McKenzie and Henry (1979) study comparing the effects of questions addressed to individual students as opposed to questions used as test-like events, it was found that test-like events reduce inattentive behavior by fifty percent. The study also showed a significant gain in achievement for the test-like events treatment group. This is not always the case in process-product studies. It seems that the positive effect of test-like events on student process variables (e.g., increased attention) does not necessarily translate into increased student achievement (Martin, 1979). There was no attempt to measure any process variables in the present study.

In attempting to reconcile the findings of past research with that of the present study, two related variables should be considered: length of treatment and delay of measuring the dependent variables. Most of the research discussed herein, with the exception of the response system studies, had very short treatments whereas the present study had a treatment lasting an entire semester. Also, many of the aforementioned studies, again with the exception of the response system studies, tested only for immediate recall whereas the present study used delayed performance measurements. As Natkin and Stahler (1969) pointed out, test-like events studies should always include a delayed performance measurement, since what works well for immediate performance may not necessarily affect delayed performance. This may offer some explanation as to why test-like events may have powerful effects in studies such as Mayer's but not so powerful for the response system studies or the present study. Likewise, what works well in a short treatment may not work well in the long term.

General Observations

The results of this study show no significant treatment effects in terms of skill level learning, no prior knowledge - treatment interaction effects, and some significant treatment effects in terms of concept/principle learning. In addressing Research Question (6), concerning the trend of results across criterion tests, it must be noted that all significant differences found were on unit examinations with none found on the comprehensive final exam. In comparing the control and both question type groups, it is not clear why a significant difference was found on concept/principle learning, Exam II, but not on Exams I and III. One explanation might simply be variation due to the different content and instruments used. No systematic content analysis was performed, but it should be noted that the focus of the material which Exam II tested was perhaps the most skill oriented (e.g., techniques of differentiation). Perhaps the both question type treatment kept those subjects in the group from overfocusing on the skill level objectives to a greater extent than on the material tested by Exams I and III.

There may be another explanation, one a bit more intriguing. While question type was controlled, question difficulty was neither controlled nor tabulated for analysis. In studying the effect of factual and application adjunct questions in prose, Watts and Anderson (1971) tabulated the performance on those questions. They reported that while the factual group correctly answered 99.6% of the adjunct questions, they answered only 52.8% of the same type questions correctly on the posttest. The two application groups correctly answered 73.2% and 62.0%

of the adjunct questions correctly and then correctly answered 70.4% and 71.6% of the same type questions on the posttest. While the focus of the study was on comparing the factual and application treatments (several other measurements were reported with the application groups showing an overall superiority), it did raise the issue of question difficulty. No cause and effect statements can be made with respect to question difficulty but it's interesting to note the gap in performance between the factual adjunct questions and the factual posttest questions and the gap (or lack of it) between the performance on the application adjunct questions and application posttest questions. The large gap in the factual scores suggests that minimal processing was required to answer the adjunct questions with minimal retention the result (Hamilton, 1985). Watts and Anderson suggest that the facilitation of learning may not be so much a function of question type as it is question difficulty, or at least a function of both. Thus, in the present study, question difficulty is one extraneous variable which might lead to a discrepancy in comparing the results on Exam II with Exams I and III.

Limitations

Fey (1980) discussed doctoral dissertation research in which the principal investigator taught one class by an experimental method and another by a traditional method. He stated that "The repeated flaws in most of the studies read like a litany of pitfalls in experimental research." A few pitfalls, unfortunately, are contained in the present study. Perhaps the most critical one is the lack of random assignment of the students to treatment groups. An attempt to equate treatment groups, based on a prior knowledge measurement, was done as outlined in Chapter III rather than attempting to correct initial differences using an ANCOVA. The ANCOVA design is commonly used in classroom experimental research but, used correctly, has randomly assigned subjects. With the additional assumptions made on an ANCOVA (e.g., no error in the covariate measurement) the present researcher felt little was to be gained by using it.

With the lack of true random assignment of students and with those students being part of intact groups, a legitimate experimental unit would be the class rather than the individual student. With three intact classes available, using this experimental unit was not possible for the design of this study.

One problem faced by all classroom researchers is that of subject mortality. Only those students who took all four criterion tests were considered for data analysis. Nine students withdrew (or simply failed to complete the course) from the control section, four from the skill only section and one from the both question type section. This problem added to the difficulty of analyzing the data and interpreting the results of this study.

One of Fey's criticisms concerned experimenter bias. Since the present researcher was the instructor of the treatment groups, this may have led to some bias.

One final limitation should be mentioned. All instruments used in this study were designed by the present researcher. The use of local instruments, while not threatening the internal validity of the study, may limit its generalizability.

The interpretation of the results of this study should be done with caution in light of the above limitations.

Conclusions

With the limitations mentioned above and the lack of clarity in the differences found among the treatment effects, it would be premature at this point to recommend the adoption of a specific test-like events treatment for normal classroom use. This study does, however, lend insight to some important questions that, when answered, could help determine when test-like events increase learning during mathematics lecture.

Clearly, the skill level test-like events used exclusively were not advantageous with respect to achievement, with slight evidence that they may even inhibit concept/principle learning. The both question type treatment was, at worst, equivalent in learning outcomes to the control treatment, with some evidence that it may increase concept/principle learning. This study raises the question of what the effect of a concept/principle only treatment would have on learning. However, it might be best to address another problem faced in this and other test-like events studies before tackling the above question. The problem is that of how best to categorize questions. Some studies (e.g., Clute, 1983) classified question type by low and high cognitive levels with others (e.g., Mayer, 1975; Wunderlich and Carry, 1974) using three or more levels. The skill and concept/principle categories used in this study are not meant to represent two distinct low and high levels of cognitive processing. For example, stating a definition would best be considered simple recall or memory level, whereas applying two or more principles in a nonalgorithmic situation might be considered either an application or a synthesis level question. However, both questions are classified concept/principle in this study. Clear rationale of question types used should be offered before future test-like events studies are conducted. In developing such a rationale, the findings of this study, as well as the theoretical framework, suggest the investigation of question types to which students might not normally attend (i.e., it doesn't look like a "normal" homework or test problem).

Tied to the issue of question type is that of question difficulty. Considered together, further investigation is needed to clarify what kind of test-like events induce more than minimal, short-term memory level processing. Past research (e.g., Watts and Anderson, 1971) suggests that there is probably an upper and lower bound on the difficulty level of test-like events if they are to be effective in facilitating learning that affects long-term memory.

Even with the most liberal interpretations of this study, one should not be tempted to generalize the results beyond the university Calculus I student. In terms of motivation, interest, study skills, intellectual development, and attention span, the typical Calculus I student is probably atypical when compared to other populations of mathematics students. Test-like events that may work well for one population may not work well for another. Likewise, the results of this study have implications only for the content of Calculus I. Future research should address not only various student populations and a variety of mathematical content but different aptitudes that may interact with the treatment as well.

Many of the above questions raised for future research could be addressed by well designed experimental studies, both in the normal classroom and in laboratory settings. However, there may be much to gain by also conducting well executed qualitative field studies. By attempting to sample the thoughts and reflections of students through interviews and self-reports, as well as observing students' on-task behavior during test-like events, perhaps many questions could be answered concerning what types of test-like events are facilitative and what frequency of test-like events is optimal.

In adjunct question in text research, highly complex experimental designs are normally required to investigate the type and direction of processing (i.e., selective or nonselective; forward or backward). In the normal classroom setting, such designs are often not feasible. Thus, in studying the attention processes of students in the classroom, qualitative research may be a suitable alternative.

BIBLIOGRAPHY

Anastasi, A. (1968). Psychological Testing (3rd ed.). London: The MacMillan Company.

Anderson, R. C. (1970). Control of student mediating during verbal learning and instruction. Review of Educational Research, 40, 349-369.

Andre, T., Mueller, C., Womack, S., Smid, K., and Tuttle, M. (1980). Adjunct application questions facilitate learning, or do they? Journal of Educational Psychology, 72, 533-543.

Beach, V. (1974). New response system: Stimulating interest and feedback. Community and Junior College Journal, 44 (7), 31-32.

Begle, E. G. (1979). Critical Variables in Mathematics Education. Washington D. C.: Mathematical Association of America and The National Council of Teachers of Mathematics.

Berndt, K. L. (1971). The lecturer at bay. The Australian Mathematics Teacher, 27, 121-126.

Bessler, W. C. and Nisbet, J. J. (1971). The use of an electronic response system in teaching biology. Science Education, 55, 275-284.

Boker, J. R. (1974). Immediate and delayed retention effects of interspersing questions in written instructional passages. Journal of Educational Psychology, 66, 96-98.

Bovy, R. C. (1981). Successful instructional methods: A cognitive information processing approach. Educational Communication and Technology Journal, 29, 203-217.

Brown, J. D. (1972). An evaluation of the Spitz student response system in teaching a course in logical and mathematical concepts. Journal of Experimental Education, 40, 12-20.

Bruning, R. H. (1968). Effects of review and testlike events within the learning of prose materials. Journal of Educational Psychology, 59, 16-19.

Cameron, P. and Giuntoli, D. (1972). Consciousness sampling in the college classroom or is anybody listening? Journal of Educational Psychology, 59, 16-19.

Casonvoa, J. (1971). An instructional experiment in organic chemistry. Journal of Chemical Education, 48, 453-455.

Chu, Y. (1972). Study and evaluation of the student response system in undergraduate instruction at Skidmore College. ERIC document ED076135.

Clute, P. (1984). Mathematics anxiety, instructional method, and achievement in a survey course in college mathematics. Journal for Research in Mathematics Education, 15, 50-58.

Cook, T. D. and Campbell, D. T. (1979). Quasi-experimentation. Chicago: Rand McNally College Publishing Company, 175-176.

Dayton, D. K. (1977). Inserted post-questions and learning from slide-tape presentations: Implications of the mathemagenic hypothesis. Audiovisual Communication Review. 25, 125-146.

Evertson, C. M., Emmer, E. T. and Brophy, J. E. (1980). Prediction of effective teaching in junior high mathematics classrooms. Journal for Research in Mathematics Education, 11, 167-168.

Fey, J. (1980). Mathematics education research on curriculum and instruction. In R. J. Shumway (Ed.), Research in Mathematics Education. Reston, Virginia: National Council of Teachers of Mathematics.

Frase, L. T. (1970). Boundary conditions for mathemagenic behaviors. Review of Educational Research, 40, 337-347.

Frase, L. T., Patrick, E. M., and Schumer, H. (1970). Effect of question position and frequency upon learning from text under different levels of incentive. Journal of Educational Psychology, 61, 52-56.

Friedman, T. L. (1976). Cognitive emphasis of geometry teachers' questions. Journal for Research in Mathematics Education, 7, 259-263.

Gall, M. (1970). The use of questions in teaching. Review of Educational Research, 40, 707-721.

Gall, M. (1984). Synthesis of research on teachers' questioning. Educational Leadership, 41, 40-47.

Gillett, P. (1984). Calculus and Analytic Geometry (2nd ed.). Lexington, Mass.: D. C. Heath and Company.

Good, T. L. (1981). Teacher expectations and student perceptions: A decade of research. Educational Leadership, 38, 415-422.

Good, T. L. and Brophy, J. E. (1978). Looking in Classrooms. New York: Harper and Row Publishers.

Hamilton, R. J. (1985). A framework for the evaluation of the effectiveness of adjunct questions and objectives. Review of Educational Research, 55, 47-85.

Hays, W. L. (1981). Statistics (3rd ed.). New York: Holt, Rinehart and Winston.

Henson, K. T. (1979). Questioning as a mode of instruction. The Clearinghouse, 53, 14-16.

Hilgard, E. R. (1964). A perspective on the relationship between learning theory and educational practices. In E. R. Hilgard (Ed.) Theories of Learning and Instruction, Chicago: National Society for the Study of Education, 402-415.

Kuehls, E. A. (1976). Effect of interspersed questions on learning from mathematics text. Journal for Research in Mathematics Education, 7, 172-175.

Kulvany, R. W. (1977). Feedback in written instruction. Review of Educational Research, 47, 211-232.

Ladas, H. (1973). The mathemagenic effects of factual review questions on the learning of incidental information: A critical review. Review of Educational Research, 43, 71-82.

Lester, F. K. (1980). Research on mathematical problem solving. In R. J. Shumway (Ed.), Research in Mathematics Education (pp. 286-323). Reston, Virginia: National Council of Teachers of Mathematics.

Levie, W. H. and Dickie, K. E. (1973). The analysis and application of media. In R. Travers (Ed.), Second Handbook of Research on Teaching (pp. 858-882). Chicago: Rand McNally College Publishing Company.

Levine, S. (1953). The role of motivation in the effects of "active review" on learning from factual film. (Abstract) The American Psychologist, 8, 388-389.

Lindsay, P. H. and Norman, D. A. (1977). Human Information Processing, New York: Academic Press.

Littauer, R. (1972). Instructional implications of a low-cost electronic student response system. Educational Technology, 12 (10), 69-71.

Lloyd, D. H. (1967). Communication in the university lecture. Staff Journal, University of Reading, 14-22.

Martin, J. (1979). Effects of teacher higher-order questions on student process and product variables in a single-classroom study. The Journal of Educational Research, 72, 183-187.

Mayer, R. E. (1975). Forward transfer of different reading strategies evoked by test-like events in mathematics text. Journal of Educational Psychology, 67, 165-169.

Mayer, R. E. (1978). Effects of prior testlike events and meaningfulness of information on numeric and comparative reasoning. Journal of Educational Psychology, 70, 29-38.

McKenzie, G. R. (1979). Effects of questions and test-like events on achievement and on-task behavior in a classroom concept learning presentation. Journal of Educational Research, 72, 348-351.

McKenzie, G. R. and Henry, M. (1979). Effects of testlike events on on-task behavior, test anxiety, and achievement in a classroom rule-learning task. Journal of Educational Psychology, 74, 370-374.

Michael, D. M. and Maccoby, N. (1953). Factors influencing verbal learning from films under varying conditions of audience participation. Journal of Experimental Psychology, 46, 411-418.

Monter, K. L. (1970). Feedback, an aid to teaching. Audiovisual Instruction, 15 (5), 89-90.

Muller, R. (1966). Student responses in lecture and instruction. Audiovisual Instruction, 11 (6), 95-95.

Natkin, G. and Stahler, E. (1969). The effects of adjunct questions on short and long-term recall of prose materials. American Educational Research Journal, 6, 425-432.

Neter, J. and Wasserman, W. (1974). Applied Linear Statistical Models. Homeland, Illinois: Richard D. Irwin Inc.

Nisbet, J. J. and Olsen, R. W. (1973). The ERAT system. American Biology Teacher, 35, 385-388.

Osterman, D. N. (1982). Classroom lecture management: Increasing individual involvement and learning in the lecture style. Journal of College Science Teaching, 12, 22-23.

Reynolds, R. E. and Anderson, R. C. (1982). Influence of questions on the allocation of attention during reading. Journal of Educational Psychology, 74, 623-632.

Reynolds, R. E., Standiford, S. N. and Anderson, R. C. (1979). Distribution of reading time when questions are asked about a restricted category of text information. Journal of Educational Psychology, 71, 183-190.

Rickards, J. P. (1979). Adjunct postquestions in text: A critical review of methods and processes. Review of Educational Research, 49, 181-196.

Rothkopf, E. Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. American Educational Research Journal, 3, 241-249.

Rothkopf, E. Z. (1968). Two scientific approaches to the management of instruction. In R. M. Gagne and W. J. Gephart (Eds.), Learning Research and School Subjects, Itasca, Illinois: F. E. Peacock.

Rothkopf, E. Z. (1970). The concept of mathemagenic activities. Review of Educational Research, 40, 325-337.

Rothkopf, E. Z. and Bloom, R. D. (1974). Effects of interpersonal interaction on the instructional value of adjunct questions in learning from written material. Journal of Educational Psychology, 61, 417-422.

Rowe, M. B. (1974). Wait-time and rewards as instructional variables, their influence on language, logic, fate control. Journal of Research in Science Teaching, 11 (2), 81-94.

Rubin, S. (1970). Evaluation of an anonymous feedback system in college classes. Paper presented at the annual meeting of the American Educational Research Association, Minneapolis, Minnesota.

Ryan, F. L. (1974). The effects on social studies achievement of multiple student responding to different levels of questioning. Journal of Experimental Education, 42 (4), 71-75.

SAS Institute Inc. (1982). SAS user's guide: Basics (1982 ed.). Cary, North Carolina: SAS Institute Inc.

Shavelson, R. J., Berliner, D. G., Ravitch, M. M. and Loeding, D. (1974). Effects of position and type of question on learning from prose material. Journal of Educational Psychology, 66, 40-48.

Sime, M. and Boyce, G. (1969). Overt responses, knowledge of results and learning. Programmed Learning and Educational Technology, 6, 12-19.

Sowder, L. K. (1980). Concept and principle learning. In R. J. Shumway (Ed.), Research in Mathematics Education (pp. 244-285). Reston, Virginia: National Council of Teachers of Mathematics.

Suydam, M. N. and Dessart, D. J. (1980). Skill learning. In R. J. Shymway (Ed.), Research in Mathematics Education (pp. 207-243). Reston, Virginia: National Council of Teachers of Mathematics.

Threadgill, J. (1979). The interaction of learner aptitude with types of questions accompanying a written lesson on logical implications. Journal for Research in Mathematics Education, 10, 337-346.

Vuke, G. J. (1962). Effects of inserted questions in films on developing an understanding of controlled experimentation (Abstract). Audiovisual Communication Review, 10, 140-141.

Watts, G. H. and Anderson, R. C. (1971). Effects of three types of inserted questions on learning from prose. Journal of Educational Psychology, 62, 387-394.

Whitehead, J. L. and Bassett, R. E. (1975). Providing feedback through student response systems. Audiovisual Instruction, 20 (8), 22-24.

Wilen, W. W. (1984). Implications of research on questions for the teacher educator. Journal of Research and Development in Education, 17, 31-35.

Wiseman, E. D. (1982). A study to determine the effect of both adjunct prequestions and adjunct postquestions on achievement (Doctoral dissertation, The Catholic University of America, 1982). Dissertation Abstracts International, 43A, 776-777.

Wunderlich, K. W. and Carry, L. R. (1974). The effects of questions of differing cognitive levels interspersed in mathematics text material. Paper presented at the annual meeting of the National Council of Teachers of Mathematics, Atlantic City.

APPENDIX A

Test-Like Events

1. Given $y = 1 - x^2$, sketch its graph and find its slope at $x = 1$.

2. Given $y = \sqrt{x}$, where is its graph rising? Falling? Flat?

3. Find the slope of $y = x^2 + 1$ at $(2,5)$.

4. Given $f(x) = x^2 + 2x$, find $f'(x)$.

5. State two forms of the difference quotient.

6. Given $f(x) = x/(1+x)$, find $f'(x)$.

7. Given $f(x) = x^{4/3}$, what does the graph look like near $(0,0)$?

8. Find $\lim_{x \to 2} (x-2)/(2-x)$.

9. Find $\lim_{x \to 0} (\sin x + x)/x$.

10. [The students were shown the graph of

$$f(x) = \begin{cases} x + 1 & x \neq 0 \\ 0 & x = 0 \text{ .]} \end{cases}$$

    (a) $\lim_{x \to 0} f(x) = ?$

    (b) Is f continuous at $x = 0$?

11. What do we mean by a function f being continuous at x_0 ?

12. Give an example of a function defined at $x = 0$ but not continuous at $x = 0$.

13. $\lim_{x \to 2} |x-2|/(x-2) = ?$

14. Given $f(x) = (1/6)x^6 + (3/4)x^4 - 2$, find $f'(x)$.

15. Define f continuous at x_0 .

16. True/False. If f is a polynomial function, then $\lim_{x \to c} f(x) = f(c)$.

17. Given $y = 3x^2 - \sqrt{x} - (4/x^3) + 5$, find dy/dx .

18. Find the equation of the line tangent to $f(x) = 5x^2 - 7x$ at $x = 3$.

19. Suppose $f(x) = x^n$, n is even. What does the derivative of f tell you about the rising and falling of the graph of f ?

20. What is the mathematical relationship between the "position" function of a moving object and its "velocity" function?

21. Given $y = \sqrt{x}/(2x-3)$, find dy/dx .

22. True/False. On parts (a) and (b) assume that f and g are differentiable.

    (a) $D_x(f(x) \cdot g(x)) = f'(x) \cdot g'(x)$.

    (b) $D_x(f(x) + g(x)) = D_x f(x) + D_x g(x)$.

    (c) If f is continuous at x , then f is differentiable at x .

23. Given $f(x) = 3x^2/(1-x)$, find $f'(-1)$.

24. $D_x(x^n) = ?$

25. $D_x(cx^n) = ?$

26. $D_x(f(x) + g(x)) = ?$ (Assume f , g differentiable)

27. $D_x(f(x)g(x)) = ?$ (Assume f , g differentiable)

28. $D_x(f(x)/g(x)) = ?$ (Assume f , g differentiable; $g(x) \neq 0$)

29. $D_x(\sin x) = ?$

30. $D_x(\cos x) = ?$

31. $h(x) = (1 - 2x)^{1/2}$. $h'(x) = ?$

32. $h(x) = \sin(2x)$. $h'(x) = ?$

33. $y = x/\sqrt{2-x^3}$. $dy/dx = ?$

34. $f(x) = [(x^2-2)/(x^2+1)]^5$. $f'(x) = ?$

35. Given: $f(0) = 0$, $f'(0) = 1$, $f(1) = 0$, $f'(1) = 1$,

$g(0) = 1$, $g'(0) = 1$, $g(1) = 1$, $g'(0) = 0$.

(a) Find $(f \circ g)'(1)$.

(b) Find $(g \circ f)'(1)$.

36. [Given information similar to that in problem 35.]

Find $(h \circ f \circ g)'(2)$.

37. Use implicit differentiation to find dy/dx in terms of x and y for $x^2y + y^2 = 1$.

38. Why would one use implicit differentiation?

39. True/False. If $\lim_{x \to a} f(x)$ exists, then f is continuous at a .

40. True/False. If f is discontinuous at $x = a$, then $f'(a)$ does not exist.

41. Given $f(x) = \sqrt{x-2}$, determine where f is continuous.

42. Given

$$f(x) = \begin{cases} (1/4)x & x \leq 2 \\ x & x > 2 \end{cases};$$

determine where f is continuous.

43. Given $f(x) = \sin x$, $x \in [-\pi,\pi]$, find the range.

44. Given $f(x) = 1 - x^2$, $x \in [-1,2]$, find the range.

45. True/False. If the domain of a function is not a finite closed interval, then the range is not a finite closed interval.

46. Give an example of a function with:

(a) a max, no min.

(b) a min, no max.

(c) no max, no min.

47. Given $f(x) = x^2 - 8$, find x such that $f(x) = 6$.

48. State the Mean Value Theorem.

49. Use the Mean Value Theorem to find "c" for $f(x) = x^3$ on $[-2,2]$.

50. [The students were shown a graph of a four part piecewise function.] Where is the function

(a) increasing?

(b) decreasing?

(c) nonincreasing?

(d) nondecreasing?

51. Given $f(x) = 1/x^2$, where is it increasing? Decreasing?

52. Given $f'(x) = \sqrt{x}$, $f(x) = ?$

53. Given $f'(x) = 1/x^2$, $f(x) = ?$

54. True/False. Two functions with the same derivative must be identical.

55. True/False. If f is concave up on I , then it is increasing on I .

56. True/False. If f is increasing on I , then it is concave up on I .

57. Given $f(x) = \sqrt[3]{x}$.

(a) Where is it increasing?

(b) Decreasing?

(c) Concave up?

(d) Concave down?

(e) Points of inflection?

58. Use linear approximation to estimate $(3.96)^{3/2}$.

59. Given $f(x) = x^{3/5}$, estimate the

(a) change in f when x is increased from 32 to 34 .

(b) change in f when x is decreased from 1 to 9/10 .

60. Use differentials to find dy/dx and dx/dy for $x^3 + 2x^2y = 7$.

61. Suppose that the domain of f is the interval $[a,b]$. Under what conditions can you be sure that the range is the interval $[f(a),f(b)]$?

62. The radius of a sphere is increasing at a rate of 2 inches per minute. What is the rate of change of the volume when $r = 5$?

63. $f(x) = x^3 - x^4$. Find

(a) all critical points.

(b) all extreme values.

(c) Graph f .

64. What are critical points?

65. $f(x) = 3x^{2/3} - 2x$. Find

(a) all critical points.

(b) all extreme values.

(c) Graph f .

66. True/False. If c is a critical point of f then $f(c)$ is a local extreme value of f .

67. True/False. If $f'(c) = f''(c) = 0$ then $f(c)$ cannot be an extreme value of f .

68. Find all horizontal, vertical, and oblique asymptotes for

(a) $y = 3x/(2x-5)$.

(b) $y = (x^2 + 2x - 3)/(x + 2)$.

69. A rancher intends to fence off a rectangular region along a river (which serves as a natural boundary requiring no fence). If the enclosed area is to be 1800 square yards, what is the least amount of fence needed?

70. Find a function in one variable used to solve the following problem:

A wire of given length can be used to make a circle, or a square, or can be cut into two pieces to make both a circle and a square. How much of the wire should be used for the circle if the total enclosed area is to be a maximum?

71. Given $y = x^2$, $x \in [-1,2]$, $P = \{-1,0,1,2\}$. Find $U(P)$ and $L(P)$.

72. True/False. Given partitions $P = \{0,1,3\}$, $Q = \{0,1,2,3\}$

(a) $U(P) \leq U(Q)$

(b) $L(P) \leq L(Q)$

(c) $\| P \| = 2$

(d) $\| Q \| = 3$

73. Given: $I = [-1,1]$, $\sum_{k=1}^{n} |c_k| \Delta x_k$. Write in integral notation and evaluate.

74. Do the following integrals exist?

(a) $\int_0^2 ((x^2-1)/(x-1))\, dx$.

(b) $\int_0^2 \frac{dx}{x-1}$.

75. Find the average value of $f(x) = 2 - x$, $I = [0,2]$.

76. True/False. Given f continuous on $[a,b]$, $\int_a^b f(x)\, dx = 0$.

(a) $f(x) = 0$ for all $x \in [a,b]$.

(b) $f(x) = 0$ for some $x \in [a,b]$.

(c) $\left|\int_a^b f(x)\, dx\right| = 0$

(d) $\int_a^b |f(x)|\, dx = 0$

(e) $U(P)$ is nonnegative for some partition P .

(f) $U(P)$ is positive for some partition P .

(g) $\int_a^b (f(x))^2\, dx = 0$

77. $\int_2^5 \frac{dx}{x^2} = ?$

78. $\int_0^2 (2x^2 - x + 4)\, dx = ?$

79. True/False. $D_x \int f(x)\, dx = \int (D_x f(x))\, dx$.

80. $\int_0^1 \frac{x^2}{\sqrt{x^3+1}}\, dx = ?$

81. $\int \frac{\sin\sqrt{t}}{\sqrt{t}}\, dt = ?$

82. Set up the integral needed to find the area between $y = \sqrt{x}$ and $y = (1/4)x$.

83. The region in the first quadrant bounded by $y = 4 - x^2$ and the coordinate axes is rotated about the x-axis. Find the volume of the resulting solid of revolution by using:

(a) the method of disks.

(b) the method of shells.

84. True/False. If f is differentiable in $[a,b]$, then it is integrable over $[a,b]$.

85. True/False. If $\int_a^b f$ exists, f must be bounded in the interval $[a,b]$.

86. Given $\int_0^1 f = 6$, $\int_0^2 f = 4$, $\int_0^5 f = 1$,

(a) $\int_2^5 f = ?$

(b) $\int_1^2 f = ?$

(c) $\int_5^1 f = ?$

(d) $\int_0^0 f = ?$

(e) $\int_2^0 f = ?$

87. What does the Fundamental Theorem of Integral Calculus (part 2) say?

88. Why doesn't the Mean Value Theorem apply to $f(x) = x^{2/3}$, $x \in [-1,3]$?

EXAM I

1. Define the derivative of $f(x)$ in terms of the limit of a difference quotient.

2. Use the definition of derivative to find $f'(x)$ for $f(x) = x - 2x^2$.

3. Use the definition of derivative to find $f'(2)$ for $f(x) = 1/x$.

4. The position of a moving object at time t is $s = 48t - 16t^2$.

   (a) What is the velocity of the object at time t ?

   (b) What is the velocity of the object at time $t = 1$?

   (c) What is the position of the object at time $t = 1$?

5. [The students were given the graph of a piecewise function and were asked the following.]

   (a) $\lim_{x \to -1} f(x) = ?$

   (b) $\lim_{x \to 1} f(x) = ?$

   (c) Is f continuous at $x = -1$?

   (d) Is f differentiable at $x = -1$?

   (e) Is f continuous at $x = 1$?

   (f) Is f differentiable at $x = 1$?

6. Evaluate the following limits. You may assume $\lim_{x \to 0} (\sin x)/x = 1$.

   (a) $\lim_{x \to 4} \frac{x^2-16}{x-4} = ?$

   (b) $\lim_{x \to 5} \frac{x-5}{5-x} = ?$

   (c) $\lim_{x \to 0} \frac{x}{\sin 2x} = ?$

   (d) $\lim_{x \to 0} \frac{\sin x + x}{x} = ?$

(e) $\lim_{x \to 0} \frac{1}{x} = ?$

7. Find the equation of the tangent to the curve $y = x^3 - x$ at $x = 2$.

8. Find $f'(x)$.

   (a) $f(x) = 3x^2 \sin x$.

   (b) $f(x) = 3x^4/\cos^2 5x$.

   (c) $f(x) = \sqrt[3]{1-x^2}$.

   (d) $f(x) = |3x^3 - x|$.

9. Use implicit differentiation to find dy/dx in terms of x and y :

   $y^3 + 2xy - y = 8$.

10. True/False.

   (a) A function f is continuous at c if $\lim_{x \to c} f(x) = f(c)$.

   (b) If f and g are differentiable at x then

   $D_x(f(x) + g(x)) = D_x f(x) + D_x g(x)$.

   (c) If f and g are differentiable at x then

   $D_x(cf(x)) = cD_x f(x)$, $c \in$ Reals.

   (d) If f and g are differentiable at x , then

   $D_x(f(x)g(x)) = (D_x f(x))(D_x g(x))$.

   (e) If f is continuous at x then f is differentiable at x .

   (f) If f is differentiable at x then f is continuous at x .

   (g) $\frac{d}{dx}(f \circ g \circ h)(x) = f'\left[g(h(x))\right] g'(h(x)) h'(x)$.

EXAM II

1. Determine where f is continuous for

$$f(x) = \begin{cases} -x & x < -1 \\ 1 & -1 \leq x \leq 1 \\ x-1 & x > 1 \end{cases} .$$

2. Given $f(x) = 3 - x^2$, $x \in [-2,1]$, find the range.

3. Given $f(x) = x^3$, $x_0 = 2$, use linear approximation to estimate $(1.97)^3$.

4. A spherical balloon is expanding under the influence of solar radiation. If the radius is increasing at a rate of 3 inches per minute, at what rate is the volume increasing when the radius is 4 inches? ($V = (4/3)\pi r^3$)

5. Given $f(x) = 1 - x^{2/3}$. Find all critical points, where the function increases and where it decreases, all local maxima, local minima, global maximum, global minimum. (If none, so state.)

6. Given $f(x) = x^5 - 5x^4$. Determine where f is concave up and concave down. Find all points of inflection.

7. A powerhouse is on one edge of a straight river and a factory is on the other edge, 5 miles downstream. The river is one mile wide. It costs \$100 per mile to run electric cable across the river and \$50 per mile on land. Write a function in one variable you would use to determine the minimum cost of running cable from the powerhouse to the factory. Do not solve.

8. True/False.

   (a) If the range of a function is a finite closed interval, then the function is continuous everywhere in its domain.

(b) If the range of a function is a finite closed interval, then the domain must be a finite closed interval.

(c) If a function is defined on a finite closed interval and is continuous, it has both a global max and a global min.

(d) Given a function f, with $f(a) = -3$ and $f(b) = 5$, there exists an $x \in (-3,5)$ such that $f(x) = 0$.

(e) If f is continuous in the interval $[a,b]$ and has an extreme value at a point c between a and b, then $f'(c) = 0$.

(f) If f is a polynomial function with a specified domain $[a,b]$, then there exists a $c \in (a,b)$ such that $(f(b) - f(a))/(b-a) = f'(c)$.

(g) All critical points are either local maxima or local minima.

(h) If $f''(c) = 0$, then $(c,f(c))$ is an inflection point.

(i) Given $f(x) = (3x^2 - 8)/(x-1)^2$, f has a vertical asymptote $x = 1$ and a horizontal asymptote $y = 3$.

(j) If $f(x)$ is continuous everywhere in its domain, it has at least one local extreme value.

(k) If (x_0,y_0) is a point of inflection, y_0 cannot be the global maximum value.

(l) If f is continuous on $[a,b]$ and decreasing on $[a,b]$ then the range of f is $[f(b),f(a)]$.

EXAM III

1. Given $f(x) = 4 - x^2$, $x \in [-1,2]$, $P = \{-1,1,2\}$.

   Compute $U(P)$.

2. Use the area interpretation of integral to find $\int_1^{-2} |x| \, dx$.

3. Evaluate the following integrals:

   (a) $\int_{-1}^{1} (3x^2 - 4x - 5) \, dx$

   (b) $\int_0^{\pi} (\sin x + \cos x) \, dx$

4. Find $\int_0^1 3x(x^2 - 1)^9 \, dx$. Let $u = x^2 - 1$.

5. Find the area bounded by the curves $y = x^2$ and $y = 8 - x^2$.

6. Given the region bounded by $y = x^2$, $x = 2$, and the positive x-axis, and rotated about the x-axis. Find the volume of the resulting solid of revolution using the disk method.

7. True/False.

   (a) If f is continuous on $[a,b]$ and G is an antiderivative of f on $[a,b]$ then $\int_a^b f(x) \, dx = G(b) - G(a)$.

   (b) $f(x)$ must be continuous on $[a,b]$ for $\int_a^b f(x) \, dx$ to exist.

   (c) $\int \sec^2 x \, dx = \tan x + C$

   (d) Given f , g , continuous on $[a,b]$ and $\int_a^b f(x) \, dx > \int_a^b g(x) \, dx$ then $f(x) > g(x)$ for all $x \in [a,b]$.

   (e) Given $f(x)$, $x \in [a,b]$ with range : $[3,\infty)$, then f is not integrable on $[a,b]$.

   (f) The function $f(x) = 2$ when $x \geq 0$, $f(x) = -2$ when $x < 0$ is integrable on the interval $[-1,1]$.

(g) The function $f(x) = 1/x$ when $x \neq 0$, $f(x) = 0$ when $x = 0$ is integrable on the interval $[-1,1]$.

8. True/False. Given f continuous on $[a,b]$ and $\int_a^b f(x)\,dx = 0$.

(a) $f(x) = 0$ for all $x \in [a,b]$.

(b) $\int_a^b |f(x)|\,dx = 0$.

(c) $U(P) \geqslant 0$ for all partitions P.

9. True/False. Assume all arbitrary functions are integrable.

(a) $D_x \int_a^x f(t)\,dt = f(x) - f(a)$.

(b) $\int \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2}\,dx = \frac{f(x)}{g(x)} + C$.

(c) $D_x \int f(x)\,dx = \int (D_x f(x))\,dx$ for all differentiable functions f.

(d) $\left|\int_a^b f(x)\,dx\right| = \int_a^b |f(x)|\,dx$.

(e) If $f(x) > 0$ for all $x \in [a,b]$, then $\int_a^b f(x)\,dx > 0$.

(f) If $\int_a^b f(x)\,dx > \int_a^b g(x)\,dx$ then $\int_a^b |f(x)|\,dx > \int_a^b |g(x)|\,dx$.

(g) If $P_1 = \{0,1,3\}$ and $P_2 = \{0,1,2,3\}$ then $U(P_1) \leqslant U(P_2)$ for any function f on $[0,3]$.

(h) $\int_a^b (f(x)g(x))\,dx = \left(\int_a^b f(x)\,dx\right)\left(\int_a^b g(x)\,dx\right)$.

(i) If $\int_2^5 f(x)\,dx = 7$ and $\int_4^5 f(x)\,dx = 3$ then $\int_2^4 f(x)\,dx = 4$.

(j) $\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$.

FINAL EXAM

1. Let f be a function differentiable at x . Express the derivative of $f(x)$ in terms of a limit of the difference quotient.

2. Use the limit of the difference quotient to find $f'(x)$ for $f(x) = x^2 - 3x$.

3. Find $f'(x)$.

   (a) $f(x) = 4x^3 - (1/x^2) + \sqrt[3]{x} - 3$.

   (b) $f(x) = \sin^3(5x)$

   (c) $f(x) = x^2/(x+1)$

   (d) $f(x) = |2x^2 - 1|$.

4. Let

$$f(x) = \begin{cases} 1 & x < 0 \\ -2x + 3 & 0 \leqslant x \leqslant 1 \\ x & x > 1 \end{cases} .$$

   For what value(s) of x is f discontinuous?

5. Evaluate the following limits. If they don't exist, so state.

   (a) $\lim_{x \to 2} \frac{x^2+x-6}{x-2}$.

   (b) $\lim_{x \to 0} |3x|/x$.

   (c) $\lim_{x \to 0} \frac{x-2}{2x+3}$

6. Find the equation of the tangent line for $y = x^2 + 2$ at the point $(3,11)$.

7. Use implicit differentiation to find dy/dx in terms of x and y : $y^2 + x^2 = xy$.

8. Let $f(x) = 3x^4 - 4x^3$. Find all local and extreme values.

9. Let $f(x) = x^3 - (1/2)x^2 - 2x + 1$. Find the intervals for which

f is concave up and concave down. Find any points of inflection.

10. Find the area of the region bounded by $y = x^2$ and $y = 2x$.

11. Evaluate.

(a) $\int_0^{\pi/2} (\cos x - \sin x)\, dx$.

(b) $\int x^2\sqrt{3 + x^3}\, dx$. Let $u = 3 + x^3$.

12. A stone is thrown straight upward. After t seconds its height above the ground is $s = 32t - 16t^2$. What is its velocity at $t = 2$?

13. True/False.

(a) A function f is said to be increasing on $[a,b]$ if for every two numbers $x_1, x_2 \in [a,b]$: $x_1 < x_2$ implies $f(x_1) < f(x_2)$.

(b) If $f'(c) = 0$ then $f(c)$ is a local extreme value.

(c) A function f is continuous at c if $\lim_{x \to c} f(x) = f(c)$.

(d) If f is continuous on $[-a,a]$ then $\int_{-a}^{a} f(x) = 0$.

(e) If f and g are continuous on $[a,b]$ then $\int_a^b (f(x) + g(x))\, dx = \int_a^b f(x)\, dx + \int_a^b g(x)\, dx$.

(f) If f and g are differentiable at x then $\frac{d}{dx}(f(x)g(x)) = \frac{d}{dx}(f(x))\, \frac{d}{dx}(g(x))$.

(g) If f is continuous on $[a,b]$ then $\int_a^b cf(x)\, dx = c\int_a^b f(x)\, dx$, $c \in \mathbb{R}$.

14. True/False. Assume f, g are continuous in their domains.

(a) If $f(x) < g(x)$ for all $x \in [a,b]$ then $\int_a^b f(x)\, dx < \int_a^b g(x)\, dx$.

(b) If $\int_a^b f(x)\,dx = \int_a^b g(x)\,dx$ then $f(x) = g(x)$ for all $x \in [a,b]$.

(c) If $\int_a^b f(x)\,dx = 0$ then $\int_a^b (f(x))^2\,dx = 0$.

(d) If f is increasing on its domain $[a,b]$, then the range of f is $[f(a),f(b)]$.

15. True/False.

(a) Given f continuous on $[2,4]$ with $f(2) = -1$ and $f(4) = 1$, there exists an $x \in [2,4]$ such that $f(x) = 0$.

(b) If the domain of f is a finite closed interval and f is continuous in its domain, then the range is a finite closed interval.

(c) If f is continuous in its domain $[a,b]$, then it has both a global max and a global min.

(d) If f has a global max and global min, then f is continuous on its domain.

(e) If f is continuous on $[a,b]$ then $D_x \int_a^x f(t)\,dt = f(x)$.

(f) If f is continuous on $[a,b]$ and G is an antiderivative of f on $[a,b]$ then $\int_a^b f(x)\,dx = G(b) - G(a)$.

16. Give an example of a function that meets the following conditions. If not possible, so state.

(a) Continuous at $x = 0$, but not differentiable at $x = 0$.

(b) Differentiable at $x = 0$, but not continuous at $x = 0$.

(c) Both continuous and differentiable at $x = 0$.

(d) Neither continuous nor differentiable at $x = 0$, yet 0 is in the domain.

(e) Continuous on [-1,1] but not integrable on [-1,1] .

(f) Integrable on [-1,1] but not continuous on [-1,1] .

17. For a function f , state three ways x can be considered a critical point.

18. [The students were given the graph of a function.]

(a) Where is the function concave down?

(b) For which interior point(s) does the derivative not exist?

19. Kathleen has two dogs. She wants to make two pens so that each dog will have the same size pen and as much ground space as possible. If Kathleen has 200 meters of fencing and the pens are to be rectangular, what should the dimensions be? (Part of the fencing is to be used to separate the pens.)

20. The radius of a circle is increasing at a rate of 2 feet per second. How fast is the area changing when the radius is 6 feet?

PRETEST

Sketch the graphs of the following. State the domain and range.

1. $y = -\sqrt{x}$

2. $y = -x^2 + 2x$

3. $y = x/|x|$

Solve for x :

4. $x^3 - x^2 = 2x$

5. $(x+2)/(x^3+5x-7) = 0$

6. $ax + by + c = 0$

7. $|x-1| > 3$

8. $\cos^2 x - 3\sin x + 3 = 0$, where $0 \leqslant x \leqslant 2\pi$

9. Let $f(x) = x^2 + x$. Find $f(3+h)$. Simplify.

10. Simplify $[2(x+z)^2 - 2x^2]/z$.

11. Find the point of intersection of the lines $y = 2x+3$ and $2x+3y = 0$.

12. Find an equation of the line with slope = 3 and x-intercept = 1 .

13. Rationalize the denominators:

    (a) $3/\sqrt{7}$

    (b) $5/(\sqrt{3}+2)$

14. Write as a single fraction: $1/6x - 3/(4x^2)$.

15. Solve for x : $x^2 \leqslant 4x-4$.

16. If $\cos\theta = 5/13$ and $\tan\theta < 0$, then $\sin\theta = ?$

17. Sketch the graph of $y = 2\cos x$.

18. $\lim_{x\to 0} \dfrac{2 - (3/x)}{1 + (4/x)} = ?$

19. Given $f(x) = \dfrac{x^2}{3x-1}$, find $f'(x)$.

20. What value of x minimizes $3x^4 - 4x^3$?

21. Integrate the following:

(a) $\int \sqrt{x}\, dx$

(b) $\int \left((1 + x^3)/x^2\right) dx$

APPENDIX B

TABLE 25
TEST-LIKE EVENTS
QUESTION CLASSIFICATIONS

Skill=S, Concept/Principle=C, Rater 1=R1, Rater 2=R2, Rater 3=R3

Item	R1	R2	R3	Item	R1	R2	R3	Item	R1	R2	R3
1	S	S	S	38	C	C	C	67	C	C	C
2	S	S	S	39	C	C	C	68a	S	S	S
3	S	S	S	40	C	C	C	68b	S	S	S
4	S	S	S	41	C	C	S	69	S	S	S
5	C	C	C	42	S	S	S	70	S	S	S
6	S	S	S	43	C	C	S	71a	S	S	S
7	C	S	C	44	C	C	S	71b	S	S	S
8	S	S	S	45	C	C	C	72a	S	C	C
9	S	S	S	46a	C	C	C	72b	S	C	C
10a	C	C	C	46b	C	C	C	72c	S	C	C
10b	C	C	C	46c	C	C	C	72d	S	C	C
11	C	C	C	47	S	S	S	73	S	C	S
12	C	C	C	48	C	C	C	74a	C	S	C
13	C	S	S	49	S	C	S	74b	C	S	C
14	S	S	S	50a	S	C	C	75	S	S	S
15	C	C	C	50b	S	C	C	76a	C	C	C
16	C	C	C	50c	S	C	C	76b	C	C	C
17	S	S	S	50d	S	C	C	76c	C	C	C
18	S	S	S	51	S	S	S	76d	C	C	C
19	C	C	C	52	S	S	S	76e	C	C	C
20	C	C	C	53	S	S	S	76f	C	C	C
21	S	S	S	54	C	C	C	76g	C	C	C
22a	C	C	C	55	C	C	C	77	S	S	S
22b	C	C	C	56	C	C	C	78	S	S	S
22c	C	C	C	57a	S	S	S	79	C	C	C
23	S	S	S	57b	S	S	S	80	S	S	S
24	C	C	C	57c	S	S	S	81	S	S	S
25	C	C	C	57d	S	S	S	82	S	S	S
26	C	C	C	57e	S	S	S	83a	S	S	S
27	C	C	C	58	S	S	S	83b	S	S	S
28	C	C	C	59a	S	S	S	84	C	C	C
29	C	S	C	59b	S	S	S	85	C	C	C
30	C	S	C	60	S	S	S	86a	C	S	C
31	S	S	S	61	C	C	C	86b	C	S	C
32	S	S	S	62	S	S	S	86c	C	S	C
33	S	S	S	63	S	S	S	86d	C	S	C
34	S	S	S	64	C	C	C	86e	C	S	C
35a	C	S	C	65a	S	S	S	87	C	C	C
35b	C	S	C	65b	S	S	S	88	C	C	C
36	C	S	C	65c	S	S	S				
37	S	S	S	66	C	C	C				

TABLE 26
EXAM I
QUESTION CLASSIFICATIONS

Skill=S, Concept/Principle=C, Rater 1=R1, Rater 2=R2, Rater 3=R3

Item	R1	R2	R3	Item	R1	R2	R3	Item	R1	R2	R3
1	C	C	C	5e	C	C	C	8c	S	S	S
2	S	S	S	5f	C	C	C	8d	S	S	S
3	S	S	S	6a	S	S	S	9	S	S	S
4a	S	S	S	6b	S	S	S	10a	C	C	C
4b	S	S	S	6c	S	S	S	10b	C	C	C
4c	S	S	S	6d	C	S	S	10c	C	C	C
5a	C	S	C	6e	S	S	S	10d	C	C	C
5b	C	S	C	7	S	S	S	10e	C	C	C
5c	C	C	C	8a	S	S	S	10f	C	C	C
5d	C	C	C	8b	S	S	S	10g	C	C	C

TABLE 27
EXAM II
QUESTION CLASSIFICATIONS

Skill=S, Concept/Principle=C, Rater 1=R1, Rater 2=R2, Rater 3=R3

Item	R1	R2	R3	Item	R1	R2	R3	Item	R1	R2	R3
1	C	C	S	8a	C	C	C	8h	C	C	C
2	S	C	S	8b	C	C	C	8i	C	C	S
3	S	S	S	8c	C	C	C	8j	C	C	C
4	S	S	S	8d	C	C	C	8k	C	C	C
5	S	S	S	8e	C	C	C	8l	C	C	C
6	S	S	S	8f	C	C	C				
7	S	S	S	8g	C	C	C				

TABLE 28
EXAM III
QUESTION CLASSIFICATIONS

Skill=S, Concept/Principle=C, Rater 1=R1, Rater 2=R2, Rater 3=R3

Item	R1	R2	R3	Item	R1	R2	R3	Item	R1	R2	R3
1	S	S	S	7c	C	C	C	9b	C	C	C
2	C	S	S	7d	C	C	C	9c	C	C	C
3a	S	S	S	7e	C	C	C	9d	C	C	C
3b	S	S	S	7f	C	C	C	9e	C	C	C
4	S	S	S	7g	C	C	C	9f	C	C	C
5	S	S	S	8a	C	C	C	9g	C	C	C
6	S	S	S	8b	C	C	C	9h	C	C	C
7a	C	C	C	8c	C	C	C	9i	C	C	C
7b	C	C	C	9a	C	C	C	9j	C	C	C

TABLE 29
FINAL EXAM
QUESTION CLASSIFICATIONS

Skill=S, Concept/Principle=C, Rater 1=R1, Rater 2=R2, Rater 3=R3

Item	R1	R2	R3	Item	R1	R2	R3	Item	R1	R2	R3
1	C	C	C	11b	S	S	S	15d	C	C	C
2	S	C	S	12	S	S	S	15e	C	C	C
3a	S	S	S	13a	C	C	C	15f	C	C	C
3b	S	S	S	13b	C	C	C	16a	C	C	C
3c	S	S	S	13c	C	C	C	16b	C	C	C
3d	S	S	S	13d	C	C	C	16c	C	C	C
4	C	C	C	13e	C	C	C	16d	C	C	C
5a	S	S	S	13f	C	C	C	16e	C	C	C
5b	S	S	S	13g	C	C	C	16f	C	C	C
5c	S	S	S	14a	C	C	C	17	C	C	C
6	S	S	S	14b	C	C	C	18a	C	C	C
7	S	S	S	14c	C	C	C	18b	C	C	C
8	S	S	S	14d	C	C	C	19	S	S	S
9	S	S	S	15a	C	C	C	20	S	S	S
10	S	S	S	15b	C	C	C				
11a	S	S	S	15c	C	C	C				

TABLE 30
CORRELATIONS OF PRETEST WITH
MEASUREMENTS OF DEPENDENT VARIABLES

Instrument	Correlation
Concept/Principle, Exam I	0.27
Skill, Exam I	0.54
Concept/Principle, Exam II	0.08
Skill, Exam II	0.57
Concept/Principle, Exam III	0.30
Skill, Exam III	0.65
Concept/Principle, Final Exam	0.45
Skill, Final Exam	0.59

TABLE 31
STUDENT EVALUATION
RESPONSE DISTRIBUTION

Compared to other O.U. courses I have taken, this course ranks:	
CONTROL SECTION	N
Far Above Average	3
Above Average	18
Average	10
Below Average	0
Far Below Average	0
SKILL ONLY SECTION	N
Far Above Average	6
Above Average	14
Average	5
Below Average	0
Far Below Average	0
BOTH QUESTION TYPE SECTION	N
Far Above Average	10
Above Average	14
Average	3
Below Average	0
Far Below Average	0

Compared to other O.U. instructors I have had, this one ranks:	
CONTROL SECTION	N
Far Above Average	13
Above Average	14
Average	4
Below Average	0
Far Below Average	0
SKILL ONLY SECTION	N
Far Above Average	15
Above Average	8
Average	2
Below Average	0
Far Below Average	0
BOTH QUESTION TYPE SECTION	N
Far Above Average	18
Above Average	9
Average	0
Below Average	0
Far Below Average	0