THE EFFECTS OF FIXED AND ASCENDING CRITERIA ON ACHIEVEMENT, ATTITUDE AND STUDY EFFICIENCY IN MASTERY LEARNING

By

James Anthony D'Albro

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Administration
and Higher Education

1980

ABSTRACT

THE EFFECTS OF FIXED AND ASCENDING CRITERIA
ON ACHIEVEMENT, ATTITUDE AND
STUDY EFFICIENCY IN MASTERY LEARNING

By

James Anthony D'Albro

Selection of the most appropriate criterion scores for criterion-referenced testing under mastery learning is uncertain because there are no procedures for sorting through the many approaches for setting such scores. Implementation strategies for nearly all the approaches are also lacking. This research was directed toward solving the problem of setting the criterion in the mastery learning strategy.

Thus, the overall purpose of this research was to identify a criterion which yielded the best achievement throughout a quarter course in greenhouse management while requiring the least amount of study time and maintaining the best student attitudes.

A different criterion level was set in each of three 50-minute mastery classes. The criteria were used in

conjunction with a mastery learning strategy. The criteria were 80% fixed, 90% fixed, and ascending (80% for the first unit and increasing 5% each unit test until 90% is reached, additional units were graded at 90% of total points). A fixed criterion was one which had the same standard applied to each of the five unit tests in the quarter.

The textual material of the mastery strategy in greenhouse management was divided into sections containing: instruction for completion, objectives for each unit, a set of review questions, and the lectures given by the instructor.

Mastery in this research was defined by three elements: instruction, grades, and testing. In order to reach mastery of a unit of study, the students had to attain the minimum criterion set for each instructional unit in a given treatment group. Whenever the criterion was not met on the first attempt of any test, the student was provided with additional instructional assistance and permitted to attempt mastery a second time. There was a total of five unit tests, each test being given at the end of a two week unit of instruction.

The control group received the same statement of objectives as the mastery groups. They were lectured on each unit of study, and were given the same test questions as the mastery groups.

The effect of the treatments was measured by a multiple choice achievement test, an attitude scale, and the total time spent on study as reported by students.

The groups taught under the mastery strategy attained a significantly higher level of achievement than the control. The study also showed that setting a higher criterion or gradually raising the criterion did not yield higher achievement than a lower criterion. The (90%) fixed criterion produced less efficient study scheduling than other criteria without any gain in achievement over a lower criterion. The ascending criterion did not produce the increase in achievement over the other criterion groups that might have been expected. Student's attitudes toward the course were also less positive when they were pushed to meet higher levels of criterion.

The findings suggested that the 80% fixed criterion could produce the best learning while maintaining the most productive student attitudes. Moreover, this criterion yielded the best student learning in the least amount of total study time over the quarter.

DEDICATION

To My Wife D'Anne
For her patience, understanding,
and constant support

ACKNOWLEDGEMENTS

The author wishes to acknowledge the assistance and advice given to this research by the following committee members:

Dr. Van C. Johnson

Dr. Max R. Raines

Dr. J. Lee Taylor

Dr. Stephen L. Yelon

The author wishes to express his sincere thanks to Dr. Stephen Yelon who spent a great deal of time editing and suggesting areas of improvement.

A special recognition is extended to Dr. Robert Smidt for his assistance on the use of statistics and computers for this research. Also, Mr. Charles Strong is recognized for his assistance on the technical writing of this dissertation.

TABLE OF CONTENTS

Page

Chapter

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Statement of the Problem

Educators are faced with the same methods of teaching and evaluating students that they used in the past. Typically a lecture hall is used to assemble a class. The instructor faces the class and begins an hour of talking about a subject. The instructor then designs a test to determine what has been remembered and/or understood. The test is graded with the intention that most students get a 'C' grade. The assignment of grades is based on the following assumptions: if students are normally distributed with respect to aptitude for some subject and all students are given exactly the same instruction, then achievement measured at the completion of the subject will be normally distributed. Aleamoni (March 1979) states that a distribution of student grades follows a normal curve so that there are 3 percent A's, 13 percent B's, 68 percent C's, 13 percent D's and 3 percent F's. But Block (1971) suggests that American education must turn away from this traditional method of teaching and evaluating. Schools must provide successful and rewarding learning experiences for

most students, not just a few. He suggests that criterion-referenced testing under the mastery strategy offers the greatest potential for students.

Mastery learning (Block, 1971) offers a powerful new approach to teaching which can provide almost all students with the successful and rewarding learning experiences now allowed to only a few. Block (1971) suggests that 75 to 90 percent of the students can reach the same high level of achievement as the top 25 percent do under traditional group-based instructional methods. Group-based instruction is teaching of a group of students at a set hour in a set room.

Further, the mastery approach includes procedures that are primarily designed for use in the group-based instructional situation, where the time allowed for learning is relatively fixed (Bloom,1968). Bloom's mastery strategy minimizes the time a student needs to learn. Therefore, most students can master the material within the calendar instructional time available.

The next section explains the essential philosophy of the mastery strategy. It also discusses some of the important features of this innovative method.

The Instructional Philosophy of Mastery Learning

Mastery learning is an instructional philosophy and an associated set of ideas about instruction. This philosophy asserts that under appropriate instructional

conditions most students can learn what they are taught (Block, 1974).

There are several procedures in the mastery learning strategy which have made the above mentioned philosophy a reasonable instructional approach (Bloom, 1974).

First, the idea of mastery does not require a normal distribution of grades from A to F. Instead, it suggests that each student should be given enough time to master the subject matter being taught. Thus, most students can get an 'A' grade.

Second, and more central to the mastery learning strategy is the use of feedback and corrective procedures. Bloom (1974) has stated that there are a variety of procedures to provide practice and feedback. They are tests, homework and workbooks. Brief diagnostic progress tests has proved to be the most useful. The test shows what the student has learned from a chapter, a unit or some other learning sequence. It also is valuable feedback to the student and the instructor on what aspects of the learning unit are weak, needing correction and further study.

Third, the diagnostic test must have some value which defines competence or mastery. The instructor can select a certain number of points correct out of a total or a percentage grade to define the needed level of competence. The benefits of declaring points or a percent is that it clearly relates achievement to the degree of mastery of what is set out to be learned. It provides a standard

measure of achievement. Consequently, students are not competing against one another. The manner by which levels of mastery have been determined is arbitrary. Herein lies a problem.

The philosophy has been converted into procedures for grading and collecting of data upon which to set a grade. As yet there is no sound basis for deciding whether or not the student can be considered a 'master'. A master is a student who has met or exceeded the criterion score set by the instructor and therefore has learned to a sufficient degree.

The evidence for a decision of mastery must be made on the basis of grades. Therefore, the problem is the setting of a certain criterion to determine grades. The criterion represents the absolute performance standard against which the sufficiency of each student's learning can be evaluated and graded (Block, 1971). This standard should indicate the specific amount of skills a student must show before he or she can be judged to have mastered the skills taught. The standard also indicates how well the student has learned. In that respect, the instructor knows exactly how much each student has learned. This is unlike a relative standard of grading which judges students in relation to others and not in relation to the course content.

An example of an absolute standard is given by Bormuth (1970) and Glaser and Nitko (1970). They state that criterion referenced tests are absolute and that these

standards define what proportion of a well defined body of content and behavior the student is expected to learn. Thus, a test with a standard of 85% suggests that a student must show competency of the content to that level.

According to Block, (1970) evaluators have ignored the problem of defining the criterion in an objective way. They have developed increasingly sophisticated data gathering instruments and procedures. But not one valid technique has been stated for defining the criterion in an objective manner. Hence, the degree of mastery of many students is being misjudged.

For example, suppose a student's learning is misjudged due to a poor criterion. The student may have to review material already learned and in so doing waste valuable study time. This time could have been spent studying material of a more advanced stage. A continuation of this defect in evaluation may eventually lead the student to a poor attitude toward a subject, a major and even to school itself. As Block (1970) states, this is most unfortunate because accurate indications of the sufficiency of a student's learning are crucial to his/her cognitive and attitudinal outcomes.

Hambleton et al. (1978) in their review of criterion levels state that the matter of the determination of criterion scores seems unclear. Further, they state that there are no procedures for sorting through the numerous approaches for determining the criterion scores in order to

select the most appropriate one for a given situation. Implementation strategies for nearly all of the approaches are also lacking. Hence, if we are to solve the problem of setting the criterion in the mastery learning strategy, a full scale directed effort must be undertaken to research the problem further. This research attempts to partially solve that problem.

In conclusion, the best available information on setting a criterion suggests that criteria are set in an arbitrary manner. It is up to an instructor to determine the level of the criteria. Therefore, this research addresses itself to the problem of the determination of a criterion score used in the mastery strategy. The research effort seeks to find an empirical basis for the criterion. The score can then be implemented when the mastery strategy is used. With this intention, the following purposes are stated.

Purpose of the Research

Instructors have selected their criteria for mastery without an explicit theory or any evidence suggesting that those chosen over others are superior in fostering student development (Block, 1970). Block (1970) goes on to say, therefore, it is entirely possible that criteria selected may not represent the best learning and attitudes as other criteria.

The research proposed has been developed to correct the major problem of setting a criterion within the mastery

strategy. Based on this research some criteria may be decided as a result of evidence.

Four steps were taken within this research to remedy the problem of setting a criterion. First, students' achievement results from criterion referenced tests were logged. These tests were diagnostic tests taken during the course.

Second, the setting of the criterion was based on a test performance which produced the highest achievement. Therefore, the instructor can evaluate what has been learned and how well it has been learned.

Third, the time to attain the criterion was within a period allocated for the course. In addition, the time to learn the content was efficient for the student.

Fourth, the attainment of high levels of achievement on criterion tests did not sacrifice the attitude of the student toward the course or major.

In an effort to correct the problem of setting a criterion, the following purposes of this research are stated:

1. It is expected that having a criterion as a goal will be of a benefit to students to achieve each unit of study in the most efficient manner. To accomplish this one may use fixed or ascending criteria to evaluate the performance of students on various units of study. Fixed criteria are absolute standards set prior to testing. They do not change from one test to the next. For example, an

80% criterion is the standard which all students are judged on each unit test. In the case of an ascending criterion, a first test may be assigned an absolute standard of 80%. Each successive test will have a new and higher percent standard by which achievement is judged.

It was hoped that by setting criteria on each test we would be able to identify levels which, when maintained throughout the learning, encourage students to learn adequately and score well on a criterion measure (Bormuth, 1969 and Block, 1970). Thus, the attainment of the criterion will indicate to the instructor that most course objectives were learned by the students. In addition, the instructor will be assured that the student has acquired a sufficient amount of course information. Therefore, the student will be judged competent in the subject.

2. The positive attitude of students may increase or decrease in response to the difficulty of a criterion. For example, a very high criterion of 90% may force students to reach that level. But, it may also cause the attitude of the student to decrease significantly. Since attitude may change, it will be the purpose of the research to identify a criterion which when used throughout the course, will produce the best student attitude toward the course.

3. Since the efficiency of learning can be interpreted as the total time it takes to learn a skill or series of skills in a unit of study, it was also the purpose of this research to manipulate the criterion or standard

used to judge students to see if students can be made to reduce their time to learn over the duration of the course.

In summary, the intention of the research was to establish a criterion which can produce the greatest efficiency and achievement without a sacrifice of the student's attitude toward the course. In addition, the aim of high achievement will also foster greater efficiency of study as manifested by time invested in the learning of content. If these objectives are attained, then a partial answer will be given as to a basis for setting the criterion under the mastery strategy.

Finally, Block (1970) suggests that if instructors can choose adequate performance levels, then educational programs might become more effective. Sound criteria make this possible. So that the research is not understated, the following statements are made to emphasize its importance.

Importance of the Study

This study was important for the following three reasons. First, the study provided a basis of setting criterion levels for a mastery program. Thus, one can base a criterion on experimental evidence. Through the implementation of an improved method of criterion selection, the student may be able to attain high levels of achievement.

Second, the proper selection of criterion is keyed closely to the attitudes of the student. There must be knowledge of a criterion which can produce a high level of

achievement without sacrificing the attitude of the student. Block (1972) states that "the attitudinal changes which do occur raises the important question of whether in pushing some students to attain very high levels of performance throughout their learning we are not, in fact, promoting their intellectual development at the expense of their feelings toward the material learned."

Lastly, the criterion selected must be attainable within the time allowed for the course. In order to meet the objective of an attainable criterion, the student must be molded into more efficient behavior. It is suggested that there may be particular criterion levels whose attainment early in the sequence will progressively increase the amount of material achieved per time later in the sequence. Students who learn under the mastery strategy, therefore, may eventually be able to achieve their required criterion level in the same amount of instructional time that should ordinarily be expected of students who learn under non-mastery conditions. Non-mastery refers to group-based instruction whereby a student is instructed and evaluated on his/her performance relative to others in the class, and where curves are established to define a spread or range of scores from 'A' to 'F'.

Research Questions

The need for this research has been stated in the above sections. In general, the research is addressed to the kinds of variations of presenting the criterion to students in order to yield the greatest achievement. In addition, in what ways can criteria be presented to students so that they maintain a high achievement and a positive attitude toward a subject? Lastly, the cognitive learning of the course content should be done with a minimum amount of time.

In an effort to provide a comprehensive answer to a selection of a criterion, the research seeks to investigate the following questions:

1. Does one criterion produce more achievement than another?

2. Does one criterion produce better student attitude than another?

3. Does one criterion produce more efficient study scheduling than another?

Research Hypotheses

From the research questions for this study, the following hypotheses were drawn. In each case, the selection of criterion was tested for its effect on achievement, attitude and total instructional time needed to learn each unit of study.

Hypotheses Regarding Achievement

The overall hypothesis of the interaction of the treatments and time for the achievement dependent variable was: There will be an interaction between treatments and time for mean achievement. The direction and the magnitude of the interaction for each treatment group is stated below,

1. The ascending criterion group will have a progressively higher score on achievement for each unit test over the period of the quarter. It is expected that early success on unit tests should motivate students to succeed later in the quarter when the criterion is at its highest level of 90%.

2. The 90% fixed criterion group will have a progressively lower score on achievement for each unit test over the period of the quarter. This group is required to reach a high level of achievement from the beginning of the quarter. We should expect early frustration in an attempt to attain this high level. There may be a loss of motivation to succeed to a high level later in the quarter if early failures are encountered.

3. The 80% fixed criterion group will have the next lowest but a moderately stable score on mean achievement for each unit test over the period of the quarter. The relative ease of reaching a low level of criterion should produce little change in achievement score from unit to unit.

4. The control group will have the lowest mean achievement score of any group for each unit test over the

period of the quarter. The students of this group are graded on a straight percent, 90%, 80%, 70%, 60% and 50% of the total score. With only one testing at each unit test, we should expect a distribution of scores of 'A' to 'F'. Furthermore, the distribution should produce an average grade of 'C'. This outcome is unlike criterion-referenced testing. Testing is done to bring most students to the highest level of achievement. Alternative learning aids are used to assist learning when mastery is not reached. Retesting is used to re-evaluate student learning. Students of the control group are not given a chance at remediation and retesting.

5. The ascending criterion referenced group, the 80% and 90% fixed criterion-referenced group will receive a higher score on a measure of achievement than students in the control class. This difference will exist because a greater number of students will attain mastery of each of the tests taken. This is so because students under the mastery teaching method are required to attain mastery or reach the prescribed criterion for each unit of study before advancing to the next unit. Therefore, more students should reach an 'A' under the criterion based grading.

6. The ascending criterion group will receive a higher mean achievement score than a fixed criterion group of 80% or 90%. The ascending criterion starts at 80% and increased by 5% in each successive unit until 90% is reached. It then remains stable at 90%. The fixed criterion (80% or

90%) remains the same for each test. The ascending criterion is predicted to be better than the 90% level because students will find it easier to achieve mastery early in the quarter. It is assumed that the early success creates a positive attitude toward the subject under this condition. To know that one can pass the tests early should motivate the students to work hard to pass tests with higher achievement levels later. Also, the maintenance of a particular high performance level later in the subject is less threatening and approached with an expectation of success. The 80% fixed criterion group should be equal to the ascending group early in the sequence. As the ascending group finds it more difficult later in the sequence, the mean difference on the achievement will become more apparent. The ascending group will score significantly higher than the 80% group since the level of achievement for the 80% group is lower.

7. The 90% fixed criterion group will get a significantly lower score on achievement than the ascending group. From the beginning, the level of achievement is set very high. Students will probably feel that this is an unreasonable expectation to meet. They will probably have much frustration in an attempt to score to a high level set for the course. Under these conditions, we can expect students to become discouraged early in the quarter. The early disappointment over failure to score properly will probably discourage students to try to score higher

in the quarter.

8. The 80% fixed criterion group will get a significantly lower achievement score than the ascending group. Since the achievement level of the 80% group is set so low for the whole term the students need not score as high as the ascending group to obtain mastery or an 'A' grade.

Finally, the 80% fixed criterion group will receive a lower mean achievement score than the 90% fixed criterion group. A 80% criterion from start to finish is set so low that students of this criterion do not have to score as high as the 90% criterion group. Therefore, the average achievement score should be much different.

Hypotheses Regarding Attitude

The overall hypothesis of the interaction of the treatments and time for the attitude dependent variable was: There will be interaction between treatments and time for mean attitude. The direction and the magnitude of the interaction for each treatment group is stated below.

1. The ascending criterion group will have the most positive mean attitude over time. This group is expected to have a progressively more positive attitude because they will attain the stated criterion for the earlier unit tests without much difficulty. The early success should motivate students to succeed later when the unit tests have become more difficult to attain. The continued success should

produce the higher positive attitude toward the course later in the quarter.

2. The 80% fixed criterion group will have the next most positive attitude over time. The relative ease of attaining criterion for each unit should produce a high positive attitude. The attitude change is expected to be moderate over time since the level of criterion can be attained without much effort.

3. The 90% fixed criterion group will have a progressively more negative attitude over the period of the quarter. While the attitude of the 90% group may start as positive as the other groups, the attitude is expected to become more negative as the criterion level continues to be difficult to attain. The author is not suggesting that it is impossible to attain very high levels of achievement but that in pushing students to do so may produce a negative response in attitude toward the course over a period of time.

4. The control group will have the most negative attitude toward the course over the period of the quarter. This group should have the most difficult time of any group in trying to succeed in a course which has a straight percentage method of grading. The students have no opportunity for remediation and retesting. We should expect the greatest frustration over difficulties in attaining a desirable score on each unit test.

5. The mastery students under criterion referenced testing will have a higher mean score on a measure of

attitude than students in the control class. One would expect a more positive attitude under mastery since those students would be given more opportunity to attain a high level of achievement. One should also expect students to be threatened less by a course which does not seek to promote a standard distribution of grades from 'A' to 'F'.

6. The ascending criterion group will have a significantly higher mean score on a measure of attitude than the 80% or 90% fixed criterion groups. Early success on the unit criterion-referenced tests of the ascending group will build confidence. In addition, it is predicted that students achieving early in the quarter are likely to report that they are learning well. Thus, it is reasonable to suggest that a positive attitude toward learning will stimulate students to master the next higher criterion level. Students will not turn away from the subject matter as they might if success on tests is hard or impossible to attain from the beginning.

7. Students in the 90% fixed criterion group will have a significantly lower mean score on a measure of attitude than the ascending group. The score which this group must attain from the beginning is set very high. There will probably be a great deal of frustration in trying to reach such a high level of achievement on unit tests. The frustration will probably be increased by the restudy and retesting that must be done in an attempt to succeed on the achievement tests. Because of this, we can expect students

to become much more negative in attitude earlier in the quarter. Since the high standard is fixed until the end of term, the attitude will remain negative.

8. Students in the 80% fixed criterion group will have a significantly lower score on the measure of attitude than the ascending group. The ease at which the 80% group can achieve each unit test will probably cause no significant change in attitude toward the course. The ascending group will also have success on the same unit test as the 80% group when the criterion is low. But, the ascending group is faced with more difficult levels of criterion late in the quarter. Therefore, the early success should motivate students to try harder to succeed later. As the ascending group continues to achieve on its tests, a more positive attitude should be noticed.

Lastly, the 80% fixed criterion group will have a significantly higher mean score on the measure of attitude than the 90% fixed criterion group. While one might be able to push students to attain a very high level of performance throughout their learning, students of the 90% group may develop a negative attitude toward the material learned. This is why we should expect significantly higher positive attitude at the lower criterion.

Hypotheses Regarding Time Spent On Instruction

The overall hypothesis of the interaction of the treatments and the time for the time spent on studies dependent variable was: There will be an interaction between

treatments and time for the mean time spent on studies. The direction and magnitude of the interaction for each treatment group is stated below.

1. The ascending criterion group will spend progressively less time per unit on studies over the period of the quarter. In the beginning of the course, students of the ascending group will have an opportunity to adjust to the course when the criterion is low. Once they have established themselves under the lower criteria, the study time of students will be less and less for each successive unit. It is suggested that students who are successful early in the quarter will not find it necessary to over-study to ensure that learning is complete.

2. The 80% fixed criterion group will spend an equal amount of time per unit on studies over the period of the quarter. A criterion of 80% is easy to reach. Once students realize that it takes little time and effort to obtain the desired score for each unit, they will spend an equal amount of time on their study of each unit to assure success.

3. The 90% fixed criterion group will spend progressively more time per unit on studies over the period of the quarter. A Criterion of 90% for each unit test is a very difficult standard to reach. If students are to assure themselves of reaching the desired criterion, a greater amount of time must be spent on studies. Any failure to reach the stated criterion level for a unit test will

indicate to the student that one must study more for the next unit test to achieve a criterion of 90%. The students will tend to add much more study time in an attempt to make sure that they have learned. Therefore, this pattern of study habit is very inefficient.

4. The control group will spend the most time per unit on studies over the period of the quarter. The opportunity to do very well in a course will depend upon performance on each test. The understanding of the content must be complete on the first attempt of each test. There is no chance for remediation and retesting. In an effort to be as complete as possible on the understanding of the course content, the students of the control group will spend a great deal of time in learning. This situation will produce inefficient study scheduling by students.

5. The criterion-referenced groups will spend significantly less mean time on study than the control group. The time spent on study can be considered a measure of study efficiency. When there is less study time spent to master a particular unit, the time spent on study is considered more efficient. There may be a particular criterion level whose attainment early in the sequence of study will progressively increase the amount of material achieved per unit time later in the course. Thus, students learning under criterion-referenced testing may be spending less time in study for tests. The students of the control group should be expected to spend a constant amount of time

through their learning. Consequently, the time spent on study will be greater for the control group than the criterion-referenced groups.

6. The ascending criterion group will spend significantly less mean time on study than the 80% or 90% fixed criterion groups. Students of the ascending group will have an opportunity to adjust to the course early in the quarter. Once they have established themselves under the lower criterion, their study time will be less and less for each successive unit. In further support of the hypothesis, it is suggested that students who find success early in the course will not find it necessary to over-study to make sure that they have learned. The overall effect will be to shape the student into an efficient pattern of studying. Therefore, it is predicted that the ascending criterion offers the greatest opportunity to provide students with greater efficiency of study. This is accomplished by a gradual incline to a more difficult criterion.

7. The 90% fixed criterion group will spend significantly more mean time on study than the ascending group. Students in the group with 90% fixed criterion will have a very high level of achievement. Therefore, it can be expected that there will be more time spent on initial learning. Additional time will also be needed for re-study and re-take tests. Overall, the time to learn to an adequate level will be greater than the ascending group.

8. The 80% fixed criterion group will spend significantly more mean time on instruction than the ascending group. Since the level of achievement for the 80% group is fixed at 80% of the total points, we can expect students to spend the same amount of study time for each unit during the quarter. The ascending group is expected to decrease in study time over the same quarter of instruction as the 80% fixed group. Therefore, the mean score on study time will be greater for the 80% fixed group.

Overview of Literature Survey

The next chapter will review the literature of mastery as it specifically relates to research on mastery learning and to the criterion setting procedures. Since this chapter has involved attitudes and the study time of the student, it is relevant to explore the literature of mastery learning with regard to these subjects. This will be done to develop background of data for specific criterion setting research procedures which produce best learning in a relatively short amount of time without sacrificing the attitude of the student.

CHAPTER II

LITERATURE SURVEY

Introduction

There are many innovative possibilities to foster the learning process. Mastery learning is a specific method of the general mastery strategy which can be used to implement this process. The general mastery strategy is defined by two essential features. One, the course content is segmented into a number of relatively short, self contained units. Students are tested on each unit.

Second, students are expected and required to meet a predetermined criterion or level of mastery before progressing to the next unit and its test.

The basic assumption of the mastery approach is that almost all students can and will learn. To meet the assumption, a set of procedures have been established. The first is that mastery entails the formulation of a set of instructional objectives that all students are expected to achieve to a particular mastery performance standard. The second procedure is the breakdown of a course into a sequence of smaller learning units where each unit typically covers several course objectives.

The third procedure is the construction of brief progress tests called formative evaluation instruments for all learning units. These tests are typically ungraded, but in some cases they may be used as the basis for the final grade. The resultant grade indicates whether the student has or has not achieved the course objectives to the appropriate level.

The final procedure is the preparation of alternative learning materials for students who have not attained mastery of the objectives of the unit. These alternatives teach the objectives in a way different than the teacher's lecture presentation.

The procedures used in this thesis were those reviewed by Block (1971). Briefly, Block reports that a subject is chosen and broken down in a specified number of units. Preferably, the subject is one requiring convergent thinking; that is, it has a definite body of knowledge upon which a group of experts can agree. Objectives are specified in a behavioral sense so students know what is expected. Ideally, the units of study build on one another. In some cases, courses may not be in a hierarchical order but are broken into units by subtopic. The students are asked to master each unit of study at a specific criterion level. The grading is, therefore, absolute in that it depends upon a level of attainment of criterion and not the class average or a curve generated from relative groups of students. When students do not reach mastery, a wide range

of procedures is initiated to help students study the same unit material and correct deficiencies. Students are then allowed to retake a unit test for mastery. The various unit tests represent formative tests; that is, tests which are not used for grade but tests which are used to inform the student of deficient areas. The summative test is used at the end of the course to put together all that has been mastered. This is the test for a grade. Bloom (1971) states that this method of learning for mastery has allowed up to 90% of the students in a particular class to achieve an 'A' grade.

Since this thesis centered around mastery learning and the use of criteria, the concept of mastery learning was reviewed in some detail.

Research Regarding Mastery Learning

Block (1971) reviewed the results from approximately 40 major studies on mastery learning. All these studies have been done under actual school conditions. They have involved all levels of education and in subjects ranging from arithmetic to philosophy to physics. Block states that these major studies have shown that 90 percent of the mastery learning students have achieved as well as 20 percent of the non-mastery learning students. Several other studies not reviewed by Block (1971) are reported below in this review.

In 1968, Amthor compared two classes of a course in descriptive geometry at the college level. Both classes

were presented with identical instruction but differed in the type of evaluation or learning strategy used. One class was taught the content of the course in lecture. They were tested on the content one time and awarded letter grades 'A' through 'F'. The students of the other class were taught under a mastery learning strategy. This strategy was explained earlier in this chapter. The results of Amthor's study were reported in terms of the number of students who received a grade of 'A' in each of the classes. The results show that 23 of the 29 students (about 80%) received a grade of 'A' for the mastery learning treatment while only 11 of the 63 (17.46%) received an 'A' in the 'A' through 'F' non-mastery graded system. Foth (1973) reports that an improved version of his mastery learning program in soil science at Michigan State University produced a grade of 'B' or better for 90% of the students; 70% achieved a grade of 'A'. In general, research by Foth found that between 70% and 80% of students received an 'A' instead of the 95% proposed by Bloom and Block. In further support of mastery learning, Wentling (1973) finds that high school students enrolled in General Automobile Mechanics obtain significantly higher mean achievement scores for both immediate achievement (test a day later) and retention (same test given three weeks later). A study conducted by Johnson, Gnagey and Chesbro (1970) contradicts the research of Foth, Amthor and Wentling. They used the mastery method whereby students were tested over the materials covered in lectures, texts

and outside readings. One group had to make a score of 80% on weekly quizzes for mastery. They were required to retest if unsuccessful until they passed. A second group was given the same four 60-item unit exams and a comprehensive final examination and assigned letter grades on the first try. A third group received no weekly test but spent the time discussing the material. None of the groups showed any increase in learning as reflected by examinations covering the material. The students were alike in their learning. The research of Johnson et al. was the only research which contradicted the positive results of mastery learning. Nevertheless, the evidence is overwhelmingly weighted in a positive direction for improved achievement under the mastery learning procedure when all of its aspects are used to teach a course.

Strategies Used To Set Criteria

The literature of the past 20 years has not reported much research on the basis for setting of criterion levels. Instead, it has produced a controversy on the validity of setting criteria.

The controversy has centered around Ebel's (1971) objection on the general meaningfulness of criteria of achievement. He states that criteria must not represent the interests, values and standards of just one teacher, but they usually do. This is true because teachers have not taken the time to come to a consensus about criteria. Therefore, according to Ebel, they lack validity and useful

meaning. Block's (1971) rebuttal to Ebel is not strong and direct. Instead, he contends that the setting of an absolute level insures that each student completes his/her learning before advancing to new information. How high a level of achievement or what knowledge is to be acquired is not answered.

With the exception of experimental papers by Block (1970) and Carlson and Minke (1975), much of the rationale used to set criterion levels for mastery learning has been subjective. In this regard Bloom (1971) states that a necessary condition for mastery is the setting of absolute performance standards. Block (1971, 1974) remarks that there are no hard and fast objective rules for setting criteria. But criteria must be set to use as the basis for grades in order to reflect attainment of those standards.

One broad suggestion regarding a strategy used to set criteria (Bloom, 1971, Block, 1971 and Millman, 1973) is to set realistic performance standards for each school or group in cooperation with teachers and administrators. The teachers and administrators would inspect test items to determine the minimum number of items that students must answer correctly in order to be considered in a "mastery state." A variation of this suggestion is proposed by Millman (1973). Test items are sorted into meaningful clusters. The clusters may correspond to the objectives of the course. Experts in the field determine the criterion score for each cluster of items. "Mastery status" could

be assumed for students whose test performance on test items in each cluster met or exceeded the corresponding criterion score.

Another educational approach for setting the criterion is Millman's (1973) approach. Millman (1973) suggests two procedures. One deals with setting the criterion so that a predetermined percentage of a group of students pass. This procedure is inconsistent with the philosophy of mastery. The philosophy asserts that students should be encouraged to achieve optimum learning of the stated course objectives. A second procedure is to administer a test to students who have already mastered the material. The criterion is chosen as the raw score corresponding to a chosen percentile score. Hambleton et al. (1978) state that this procedure has its limitation but they do not state why it is limited.

A third approach for setting criteria is that grades for the following year might be based on grading standards arrived at the previous year if parallel examinations are used. Specifically, Block (1971) states that scores which earned students learning under non-mastery condition 'A's' and 'B's' might be useful mastery grading standards. Based on Block's suggestion, Hapkiewicz and Foth (1973) have reported that scores which earn students 'A's' or 'B's' in a previous term when grades were assigned on a curve were specified as the standard for students in mastery learning courses. A scale was developed from previous course grades

in Soil Science 210 at Michigan State University (Hapkiewicz and Foth, 1973). The scale was more rigorous than most previous scales used since no one received a grade point average of 4.0 with less than 88%; whereas some students received a 4.0 with only 84% when grades were based on a curve under the non-mastery system.

A fourth approach is suggested by Hambleton et al. (1978). They claim that, in general, criterion scores probably should be based on psychological and educational considerations, but in some instances statistical considerations can be brought to bear on the problem of setting a criterion score. Several statistical procedures, which they reviewed, are stated in the following paragraphs.

Huynh and Perney (in press-see Hambleton et al. 1978) suggest a method of estimating criterion scores. Test performance data for a group of students on a series of unit tests plus test scores from a "referral task" are needed to start their algorithm for criterion score and domain score estimation. On the basis of an initial classification of students into mastery states, determined by data obtained from the referral task, a score and domain scores for the last unit in the sequence can be obtained. The criterion score and mastery determination from the last unit will then serve as the "referral task" data for the second to last unit. The process is continued until scores and domain score estimates are available for all students on each of the unit tests.

According to Hambleton et al. (1978) the practical value of Huynh and Perneys' method of estimating criterion scores is unknown. The method of Huynh and Perney appears to have several problems. It assumes all items in a unit test to have equal difficulty. It requires the existence of an independent measure of performance, to which they referred in their work as a "referral task." There must be the proper sequencing of units. Also, there is a subjective assignment of students into mastery states based on the referral task.

Berk (1976) proposes a relatively simple procedure for selecting a criterion score. The method requires the selection of instructed and uninstructed groups of students. Instructed students are those who have received "effective" instruction on an objective to be assessed. Effective instruction involves a qualitative judgement about the mastery of an objective by students. Uninstructed students are those who have not received instruction on an objective. They are also tested to see if they have mastered the objective.

Generally, the distribution of instructed and uninstructed student scores, ranging from zero to i, where i is the number of items on the test, can be divided by a series of criterion scores into two general categories; masters and non-masters. According to Berk, a criterion which produces the greatest frequency count of students at or above the criterion identifies the groups generally

considered masters. Those below the criterion are considered non-masters. Since it is assumed that the students of the instructed group are 'true masters', these students are specifically put into two classes: True masters (TM), and false non-masters (FN). Similarly, the students in the uninstructed group are classified as false masters (FM) and true non-masters (TN). The classification just mentioned is expressed in a box form below.

Predictor Classification (Cutting Score)	Criterion Classification: Instructed (I)	Criterion Classification: Uninstructed (U)
Predicted Masters (PM=TM + FM)	True Masters (TM)	Type II Error False Masters (FM)
Predicted Nonmasters (PN=FN + TN)	Type I Error False Nonmasters (FN)	True Nonmasters (TN)
	Masters (M=TM + FN)	Nonmasters (N=FM + TN)

For clarity, the cells of the above box are identified by a classification term for the instructed and uninstructed students. In practice, the probability of scoring on the test is placed within each cell of the box. The probabilities of the four classifications can be obtained by simply expressing the frequency count of the number of students placed within each classification as proportions of the total sample. For example, the proportion of true

masters equals the true masters divided by the total number of instructed and uninstructed students in a sample. The optimum criterion score is the one that maximizes the proportion of correct classification. The proportion of correct classification for a particular criterion is equal to the proportion of students in the instructed group assigned to a mastery state (TM) plus the proportion of students in the uninstructed group assigned to a non-mastery state (TN). We may then assign the optimum criterion to a particular course.

Factors to Consider For Setting Criteria

There are several factors to consider for creating criteria. These factors are proposed by Millman (1973). Recently, they have been reviewed by Hambleton et al. (1978). The factors are:

1. educational consequences
2. psychological and financial costs
3. errors caused by guessing
4. item sampling

The educational consequences involve setting higher criterion scores for fundamental or prerequisite skills. Millman (1973) states that skills that are not prerequisite to others may not require criteria at all. He suggests the higher criteria are needed with prerequisite courses to make sure that students are well prepared for advanced courses. Setting the criteria too high may prove wasteful of teacher and student time and resources.

A consideration of psychological and financial costs led Millman (1973) to suggest that a low criterion score should be set when remediation costs are high. In situations with lower remediation costs or with higher costs associated with false-positive errors, (a marginal pass that is not a pass) high levels of a criterion should be considered.

Errors caused by student guessing may lead to not classifying certain students as masters. For this reason Millman (1973) states that there may have to be a correction for guessing to adjust the criterion score. How to make the correction was not treated.

The errors introduced by item sampling is a bias resulting from systematically disregarding some of the types of questions and some content in the domain of test items measuring an objective. Knowledge of this bias has not led Millman to any conclusion as to a method of correction. Perhaps, the bias can be minimized by careful consideration to test construction or a clear and concise writing of the objectives for a course.

Setting the Level of the Criterion

As a result of Block's research (1970, 1972), there may be an objective rational for establishing a criterion. His research shows that when one wants a great deal of learning, selection of high criterion (95%) is appropriate, but when there is a greater concern for the attitude of students, a lower criterion should be set. In his case, 85%

was used. Furthermore, Block states that a criterion between 85% and 95% can be chosen which gives a desirable blend of achievement and attitudinal outcomes. It is pointed out by Block that the above results must be interpreted cautiously until they can be reproduced with a much larger sample, on a longer learning sequence and in a variety of subjects.

Carlson and Minke (1975) worked with 147 students. Three consecutive 10-week night classes of Survey of Psychology, at the University of Hawaii were used in the study. Carlson and Minke tested three experimental criteria; that is, ascending from 60% to 90% and two fixed criteria of 80% and 90% for all quizzes. The students were informed that a 10-item multiple choice quiz on each unit was to be mastered at a stated criterion level. The stated criteria for final grades were based on the total number of units passed.

The description of the above research differs from the research of the author in the following ways.

1. The ascending criterion of the author started at 80% and increased by 5% until 90% was reached. The criteria were applied to each of 5 unit tests. Carlson and Minke started at 60% and increased by 10%. The criteria were not applied consecutively to each of 28 units. For example, units 2, 8 and 9 were graded at 70%.

2. The research of the author was done in the same quarter. Carlson and Minke used three consecutive quarters.

3. Each unit in the research of the author had to be passed before the student could take the next unit test. All units had to be mastered. The final grades of Carlson and Minke were based on completion of units. For example, 15 units had to be passed for a grade of A. No requirement was specified for any order of units to pass.

Carlson and Minke found that the highest criterion, 90%, produced the lowest number of high course grades and passing grades. This is contrary to Block's (1970) notion that a 95% criterion level produces the greatest learning. The best performance was shown by the 80% fixed group for final grades, passing grades per unit test and cumulative units passed per student. One should expect a high passing rate for the lower fixed criterion. Students of the 80% fixed group did not have a criterion level as high as others to master each unit. The sixty to nenety percent ascending group required fewer attempts overall to master unit quizzes than did either 80% or 90% groups. Carlson and Minke believe that the early success felt by students helped to reduce frustration which may have been felt by the other groups. The effect may be positive reinforcement to continue to perform. This was evidenced by a tendency of the 60% to 90% group to make fewer errors than the 80% group on units when the criterion was 80%. Carlson and Minke (1975) believe that the enhanced performance on tests of the 60% - 90% group was a result of 'shaping'. That is, the high level of performance later is brough about by a gradual

increase in the mastery criteria. However, the 60% - 90% groups passed fewer units on the first take when the criterion was 90% when compared to the 90% fixed group. This does not fully support the concept of shaping. Instead, it appears that students might have come to "motivational ceiling." They could have decided that a lower grade level was good enough. Carlson and Minke suggest that the ascending group may have reinforced less than optimal study habits which persisted and retarded performance on later units. This explanation supports a "motivational ceiling" effect suggested by the author.

Summary of Research Regarding Mastery Learning

In summary, achievement gains under the mastery strategy appear to be well documented in the literature. Students who learn under the mastery strategy with criterion referenced testing achieve and learn more than students in non-mastery classes.

In addition, the mastery strategy produces a greater number of grades in the 'A' and 'B' category than under other methods of instruction. The reason is that students are asked to achieve to a certain level of performance. The performance level is designated as the standard which demonstrates the best learning of course content. Attainment of that level is awarded an 'A' grade. The level of criterion is keyed to the best understanding of the objectives of the course. When students fail to reach criterion, they are asked to use alternative learning systems

(test, tutor, audiotape, etc.) to study those lecture objectives which are not clear.

There does not appear to be a concrete strategy for setting criterion scores. What is available in the literature? There are several educational and statistical approaches. Educationally, groups of experts in a field in cooperation with administrators may decide upon the relevance of a criterion to the content of a course. Additionally, the prerequisite status of a course may influence the experts and administrators on how difficult the standard should be.

Criterion scores may be set by using scores which earned students 'A's and 'B's under a non-mastery approach to teaching. This is even better if parallel examinations are used for the mastery groups.

A criterion score may be set so that a pre-determined number of students pass. This method is inconsistent with the philosophy of mastery and criterion-referenced testing. The philosophy is that students must be evaluated on their absolute performance on the stated course objectives. The instructor sets the level of achievement so that the instructor can evaluate whether or not the stated objectives have been met by the student. All students must be given the opportunity to master the course content.

Lastly, there are a few statistical procedures which may help to set criterion scores. All the procedures are based on classifying students into two categories: masters

and non-masters. One statistical approach is to subjectively assign students to the categories of mastery and non-mastery. This is based on independent measures of performance and a series of unit tests. From this data, the criterion score is determined for the next group of students. On the other hand, Berk suggests that a criterion score should be determined by using the test scores from samples of instructed and uninstructed students. A series of criterion scores are used to determine which criterion produces the greatest number of scores for students at or above the criterion. These students are masters of the test. The instructed group of masters are classified as 'true masters' while the uninstructed group is classified as 'false masters.' The instructed and uninstructed groups who do not pass the test are classified as 'false non-masters' and 'true non-masters' respectively. The optimal criterion is selected according to the estimated probabilities of correctly classifying students. The probabilities are obtained by dividing the number of scores for each classification by the total number of students. These probabilities are proportions of the total sample. The proportion of students in the instructed group assigned to mastery plus the proportion of students in the uninstructed group assigned to non-mastery equals the optimal criterion for a particular course.

Several factors should be considered before creating criterion scores. They are:

1. educational consequences
2. psychological and financial costs
3. errors caused by guessing
4. item sampling

In terms of educational consequences, Millman suggests that prerequisite courses should have criteria. The criteria should be higher than other courses. The higher standard assures that students are well prepared for advanced courses. Courses which are not prerequisite probably do not need criteria.

When psychological and financial costs are high, low criterion scores should be set. When costs are low, creating higher criteria may be more reasonable.

Millman (1973) suggests that errors caused by student guessing must be corrected by some method. Without the correction, students may not be classified as masters. No method of correction is proposed by Millman (1973).

The last factor to consider for creating criteria is item sampling. A bias may result from disregarding some of the types of questions and some content in the domain of test items measuring an objective. While the problem is recognized by Millman (1973), no suggestion on correction of the problem is proposed. The author suggests that it may be corrected by careful consideration of test construction. Also, precise writing of objectives may minimize the problem.

Research has also been directed toward setting the level of the criterion. The research reported in this paper does show that there is a criterion which produces a desirable blend of achievement of course objectives and attitude outcome. Block believes that a criterion between 85% and 95% produces this blend. On the other hand, Carlson and Minke propose that ascending the criterion in a course (start low and increase to a high) is better because it produces a student with a better attitude later in the course. A high level of performance later is brought about by a gradual increase in the mastery criteria. This is not fully supported by the research of Carlson and Minke. Students taught by a fixed criterion of 90% achieved more than the ascending group when their criterion was 90%. The researchers suggest that the ascending group may have reinforced less than optimal study habits. This persisted and retarded performance on later units.

Study Time Needed to Attain Criterion

Block (1972, 1974) reports that there is little doubt that the mastery group characteristically requires additional time and help to bring them to the particular criterion established by an instructor. Furthermore, Block (1974) states that in order to bring 80% of the students to the level of achievement attained by 20% of the students under non-mastery conditions, 10 to 20% additional out-of-class study time is needed for certain students. Wentling (1973 found that immediate

and delayed achievement was significantly higher for a fixed criterion group of 80% but that the amount of time spent on instruction was 50% greater for this group as compared to non-mastery students. Wentling's measure of time was to have all students keep a record of all time spent upon instruction and testing for each unit. Perhaps, the efficiency of the students' study habits can be increased, eventually decreasing the time spent on instruction and testing. An indication of this is noted by Block (1974). He observed that students under mastery varied in the beginning quite a lot with regard to extra time needed for mastery. As a term progressed, the students became more alike in their learning efficiency as measured by time devoted directly to the learning effort. Block (1971) and Glaser (1968) suggest that perhaps this initial difference in study time is due to aptitude levels and that these levels are less obvious when time is varied for individuals. This led Carroll (1970) to conclude that each student has a time to attain the criterion and that the time to learn is the aptitude of each student. Mastery strategies offer the needed variation for each student to come to criterion and to gain in learning efficiency.

Some literature has showed that under certain conditions mastery can be achieved successfully in shorter periods of time. In other words, as the mastery criterion level is increased, the time to master the unit decreases. Block (1972) attempted to test this relationship by assigning

different criterion levels (65%, 75%, 85%, 95%) to groups of students. These levels are the percent correct of total score needed to be considered as a pass and therefore mastery. Block measured total amount of learning time as an indication of efficiency of studies. The total time to learn included textbook learning and time spent on correction and review for each unit. It is concluded by Block that all of the mastery groups spent more learning time than the non-mastery treatment group. Also, the 75, 85 and 95 percent treatment groups spent the same total amount of learning time. In the same amount of time, the 95% group achieved more course content than the 85% group and the 85% group achieved more course content than the 75% group. This situation indicates that the 95% group learned more efficiently than the 85% and the 85% group learned more efficiently than the 75% group. These results can only be taken as tentative because Block's sample was very small, the learning sequences were short and the age of the subjects and the course matter used to make the study was very limited in scope. Several students had dropped out of several of the treatment groups which would tend to bias the results. Also, the control group and the treatment groups were in the same classroom, which may have created a competitive atmosphere among the groups, thus causing another treatment bias.

Summary of Study Time Needed to Attain Criterion

In summary, specification of ways to reduce the time spent on study and testing appears to be inconclusive in the literature. Variables reported are so difficult to control that the knowledge of their effects on time use may not be known for a long time. The motivation of the student, the quality of instruction, the prerequisite background of the students and the previous study habits of the students are a few of the factors which create confusion when investigating time to learn and efficiency of study under the mastery strategy. Additionally, the method of collection of the data adds difficulty to the problem. Some individuals allow students to report data while others observe the time used to study and/or test. Also, the kind of data collected varies. Researchers have used total time. Total time is the amount of time used to read textbooks, study for tests, do study projects, take tests, and do correction and retakes on unit tests. Others were using time spent by the student on instruction to complete a unit. In order to equate the allocation of the time on unit tasks, the latter is preferred.

Student Attitudes and Learning for Mastery

It would be best to define the subject of attitude as influenced by mastery before it is surveyed. Bloom (1971) described attitude as a general disposition to regard something in a positive or negative way. It is a feeling which attracts one toward a subject or repels one away.

Many researchers have come to a common conclusion; that is, when a student does well in a subject and more generally in school over time, he/she continues to develop a positive attitude toward the object. In a general sense, Bloom (1971) has also made these conclusions regarding attitude in school. First, if a student develops a negative (or positive) attitude toward school, it may include the subjects, the teachers and staff. It may also include the whole idea of school and school learning. Second, different amounts of failure (or success) may be needed for different students to develop this negative or positive attitude toward school. It is a matter of degree. All individuals who accumulate sufficient experiences of failure (or success) will at some point develop negative or positive attitudes toward school. Third, the degree of certainty of attitude formation is likely to be much greater for negative attitudes and repeated evidence of inadequacy. Last, other variables determine whether the school and school learning is viewed as positive and favorable, e.g., values of parents, peer group attitudes, meaningfulness of schooling for the individual's career aspirations. Bloom (1971) concludes that in order for a student to view himself in a positive way, he/she must be given opportunities for rewards. Mastery learning provides the necessary reassurance and reinforcement to assure a positive attitude.

Wooford and Willoughby (1968) measured attitude in order to predict scholastic behavior. Seventy-two students

of general psychology answered a 40-item sentence completion attitude scale which measured attitudes toward two specific factors: instructor and the course. Two general factors measured were college and life. Scholastic behavior measures were absences, tardiness and course grades. They concluded that the best predictor of this scholastic behavior is the composite attitude scores (instructor, course, college life). Of equal or greater interest was the finding that course grades were significantly related to the attitude toward the course but not significantly related to attitude toward college.

Neidt and Hedlund (1967) found that student attitudes toward a particular learning experience become progressively more closely related to achievement in the learning experience as the period of instruction progressed. In this case, attitudes remained very course specific, but as mentioned earlier, a continuation of negative or positive success could lead to a general negative or positive attitude about school.

Reports by Harris et al. (1969), Neidt and Hedlund (1967),and Sheppard and MacDermot (1970), indicated that there is a correlation between high success and high positive attitude. Harris et al. (1969) and Sheppard and MacDermot (1970) suggest that a positive attitude is a very significant asset of the mastery strategy. According to the latter authors, students are systematically led to success on units of study and a course in general.

Block (1970, 1972) suggests that attitudes may become negative when high achievement is established under mastery. Block tested attitudinal changes by using 91 eighth graders who were taught a three-unit sequence on matrix arithmetic. Sixteen students in each of four classes were assigned to mastery treatments (four students per treatment). Each treatment helped the student to reach a particular performance level, for example, to attain either 65, 75, 85, or 95 percent of the material in each unit. The percent of material attained was the student's score on diagnostic-prescriptive unit tests. Other students in each class were assigned to a non-mastery treatment. They were not required to attain any particular performance level.

Block's results indicated two important points. First, students of the mastery treatments had a significantly higher attitude score toward arithmetic than the non-mastery group.

Second, the achievement scores and attitude scores toward matrix arithmetic increased up to the 85 percent performance level. The achievement scores of 95 percent performance level also increased while scores on the 24 item attitude questionnaire showed a decline in attitude toward matrix arithmetic. Therefore, a mastery strategy which forces attainment of very high achievement scores may eventually cause a decrease in attitude toward the subject.

Summary of Literature on Student Attitudes

In summary, student attitudes are correlated to academic achievement. When students succeed in a course, they generally develop a positive attitude toward a subject and school. A different amount of success or failure is needed by each student to develop a positive or negative attitude toward school or his/her course work.

Attitude formation is likely to occur with greater certainty when failures in course work are continually encountered in school.

Student attitudes toward a particular course is closely related to course grades. Student attitudes toward a course become progressively more closely related to achievement in the course as the term advances.

When instructors set criterion levels in mastery learning, they should consider the attitudinal outcomes of students. Unreasonably high levels of performance can lead to negative attitudes toward the course. If this becomes a consistent pattern, the negative attitude of the student can extend to his/her major or even the school.

Summary of Literature Survey and Relation to Research Questions

While it is clear that gains in achievement are possible under the mastery strategy, it is not clear what criterion score determines the greatest learning. Thus, the major concern of this research is to search for a criterion score which will produce the best

learning on each test. The results of the hypothesis written on achievement will be used to determine the selection of criterion which produces relatively greater achievement. It would be a step toward setting a minimal level of performance that students should be required to maintain throughout their learning. A minimal level in this case is one which, under certain conditions, is the best learning of course content that we are able to produce. If this research hypothesis is supported, it will then partially answer the question of which criterion level produces the best learning. A consideration of setting a criterion score leads to two other dimensions mentioned in the survey of literature. They are student attitudes and the amount of time spent on studies.

It is certainly important that instructors find ways to produce the highest learning of course content. There are indications though, that influences on learning can also influence student attitude. Thus, the second major research question was: Is there a criterion score which will produce relatively positive feelings toward the subject? This question points to a relationship between achievement and attitude. The indication in the literature is that a student's achievement in a course in turn influences his/her attitude toward a subject. It seems reasonable that the student's perception of his/her learning adequacy should influence his/her academic attitudes. The student's ability to maintain particularly high criterion

levels would likely produce a positive perception of his/her learning adequacy (Block, 1971). Attitudes should contribute to learning. In turn, learning should positively reinforce a predisposition toward a certain attitude. Thus, the results of the hypotheses on attitudes may answer the question of whether or not a particular criterion can cause a relatively positive attitude toward the course.

Time spent on studies also appears to be related to achievement. The few studies on the subject indicate that maintenance of particular criterion scores has an effect on study time on a task. General statements about the effects are inconclusive because the research is conflicting. Some reports such as Block (1970) and Wentling (1973) show more time spent on instruction under criterion-referenced testing while others such as Block (1974) report decreased time spent on the task. The literature is further complicated by results which show different criterion scores producing the reduction in time spent on instruction.

Furthermore, it may be that student attitudes influence their use of time. There appears to be a "motivational ceiling" developed by students with regard to time and difficulty of criterion. It may be that students feel that a particular grade is good enough and no further effort is necessary. This is supported by the research of Carlson and Minke. Their ascending criterion (60, 70, 80, 90%) did not fully shape students to succeed at the highest criterion as compared with a group maintained at 90%

throughout the term. It is suggested by Carlson and Minke that the lower criterion in the beginning may have gotten the students off to less than efficient study habits. Later, this inefficiency was maintained instead of improving.

The inconclusive nature of the research on study time under mastery makes it a likely target for further study. The relationship between the maintenance of a particular criterion throughout learning and the time students need to learn must be studied in greater depth. The lack of evidence on study efficiency under mastery has made the following question an important part of this research. Are there criteria which help to make a student more efficient in his/her studies as the quarter progresses? In this study, efficiency is expressed as time needed to learn a unit of study. Time is defined as homework, textbook study, note study, extra reading assignments, tutoring and any other time directly spent to learn the content of a unit of study. The results of the hypothesis on study time will partially answer the question just mentioned.

CHAPTER III

PILOT STUDY

Introduction

There are three questions to be answered by this research. First, what criterion will yield the best learning?

Second, what criterion will yield the best attitude toward a course?

Lastly, what criterion will yield the best study efficiency?

In order to find an answer to the above questions, three criteria and a control were used as treatment variables. The three criteria were 80% fixed, 90% fixed and ascending. A fixed criterion remained the same for each test of the quarter. An ascending criterion increased in percent over the quarter. In this research, the ascending criterion was 80%, 85%, 90%, 90% and 90% for each test respectively. Mastery under the criterion-referenced testing is defined by a score equal to the multiplication of the criterion percent by the total possible points for each test.

The control group was treated on a straight percentage of total points. Thus, each test would have a

percent of 90%, 80%, 70%, 60% and 50%.

The dependent variables of achievement, attitude toward the course, and student study time were the measured variables used in the study.

First, the achievement score for each test was used to analyze the effect of the treatment variables. The mean achievement scores would be compared to determine group differences.

Second, the attitude scale was used to analyze the attitude of the student toward the subject as related to the treatment variables. The mean attitude score was compared to determine group differences.

Lastly, the study time of the students in each treatment group was reported by students for each unit. The time for course studies in this research was defined as total hours directly related to learning of content.

The pilot study was undertaken to assess the validity and reliability of the achievement test. Also, the attitude scale was constructed during this part of the research. The reliability of the attitude was also calculated. Lastly, the pilot was used to evaluate the objectives and the textual material to be used in the study. The pilot did not evaluate the dependent measure of study time to be used in the research because the time would be reported by students during the treatment part of the research.

Population and Sample

The population of the pilot study consisted of seventeen junior and senior state university students at the California Polytechnic State University, San Luis Obispo, California. The students were enrolled in a Greenhouse Management course during the Summer Quarter of 1976. Eighty per cent of the students were seniors. The average age of the students was 22 years. Twelve students in the pilot were transfer students. Most students can be classed as elective students. Elective students chose the course freely.

The pilot group was not informed that this was a preliminary research study. To do so may have caused students to act in an unnatural way towards the course. This is typically referred to as the Hawthorne Effect. The group might have done better on the measurement instruments because they knew that they were being studied.

Course Material and Instruction

The objectives (see Appendix A) were passed out to the group of 17 students in a Greenhouse Management course, Ornamental Horticulture 323-01. The students were informed that the objectives were related to the lectures, handouts, and assignments, and that unit tests were derived from the lectures, handouts and assignments. The lecture material or course content was primarily disseminated orally by the lecture instructor during the scheduled hours for the course.

The handouts supplemented the lectures. When desirable, students were able to use several greenhouse management texts as references.

Course Evaluation

Students were informed that grades were derived from a single administration of each of the five unit tests. Each unit test was related to a defined amount of content as represented in the objectives for each unit. The unit tests were assigned a score of 90% in order to receive an 'A' grade. Grades of 'B', 'C', 'D', 'F' are rated at 80%, 70%, 60%, 50%, respectively. Final grades were computed by an average of all unit tests taken. Averaging the scores seemed to be more of an incentive to do better on the individual unit tests than any other method. Students may have felt that they had a better chance for a higher final grade when they were tested on a smaller amount of course content.

Other Types of Study Aids Used in the Course

Students were encouraged to use the recommended texts listed in the written introduction of the course. These texts related to all of the objectives at one time or the other. There were also numerous agricultural extension bulletins made available to students as the need arose.

Lastly, the students were instructed that there were numerous human research personnel located at the state university. These people could help students in the understanding of any lecture content. Students were also reminded that their

lecture instructor was available for out of class tutoring of any lecture content.

Validity of the Achievement Test

Ebel writes that the standards for Educational and Psychological Tests and Manuals delimits three kinds of validity for tests: content validity, criterion-related validity and construct validity. Criterion related validity determines the extent to which scores of a test provide useful estimates of a student's knowledge in a subject. In order to validate the scores, one would need a test generally accepted as or known to be valid. Such a test was not available. Therefore, this kind of validity was not used.

Construct validity is the degree to which test scores measure particular psychological traits. Some of these traits are creativity, anxiety and practicality. The terms are the constructs being validated. This study did not measure any psychological trait which would rate students on a particular construct.

For the purposes of the pilot study, content validity was used. The study was interested in the extent to which the content included in the unit achievement test was a balanced and complete sampling of the knowledge, skills and understanding the instructor was attempting to develop in the course (Ebel 1973, Erickson and Wentling 1976). Therefore, the content validity was determined by comparing test content with the instructional objectives for the course (Erickson and Wentling, 1976).

The comparison was done by the author. Other individuals were asked to assess the validity of the content, but all declined. The basic argument was that the author was the one who best knows the content. This is reasonable because the course used in this study had been taught by the author six times prior to this study.

The author also had made numerous test questions in the past on course content in greenhouse management. Lastly, the industry experience as a greenhouse and personnel manager have added expertise to relating evaluative instruments to course content.

According to Ebel (1973), there is no commonly used numerical expression for content validity. It was determined by a thorough inspection of the items of the test by the author. In order to do this inspection, a table of specifications for direct assessment of student performance objectives was used. This was patterned after Erickson and Wentling (1976). The table on the next page shows the objectives by unit with a check-off system for identification of objectives which were included in the achievement test. As is shown, there was a high agreement between the objectives and the achievement test for the first four units of study. Objective 2 under the heading 'Define Management' and objective 3 under the heading 'Describe types of business ownership' have been deleted. These objectives were dropped because they were never part of the course content. Therefore, they were never written into the unit

Table 1 . Table of Specifications for direct assessment of objectives covered on five unit achievement tests in the pilot study. See Appendix A for number corresponding to objective.

Objective by Unit	Objectives Included in Achievement Test
I. Define Management	+
1	
2	
3	+
4	+
Describe types of Business Ownership	
1	+
2	+
3	
Describe different Marketing Set-Ups	
1	+
2	+
3	+
4	
Apply Market set-ups to Business	
1	+
2	+
Diagram Organization Flow Charts	
1	+
2	+
3	+
4	+
II. Identify Different Recruitment Procedures	
1	+
Describe Orientation Procedures	
1	+
2	+
3	+
Plan Training Procedures	
1	+
2	+
3	+

Table 1 . CONTINUED

Objective by Unit		Objectives Included in Achievement Test
III.	Estimate Production Peaks	
	1	+
	2	+
	3	+
	4	+
	Calculate Year Around Crop Rotations	
	1	+
	2	+
	Calculate Number of Plants/Pots	
	1	+
	2	+
	3	+
IV.	Schedule Year-Around Crops	
	1	+
	2	+
	3	+
	4	
V.	Compile Cost Estimates	
	1	
	2	
	Analyze Selected Cost	
	1	
	2	
	Define Financial Terms	
	1	+
	Compute Cost of Production	
	1	+
	2	+
	Describe Profitability	
	1	+

tests. Objective 4 under the heading 'Describe different marketing set-ups' was covered in future administrations of the achievement tests.

Unit 5 of the pilot study had the greatest deficiency of coverage of objectives by the achievement test. For this reason, objectives were rewritten to match the course content. The complete set of rewritten objectives is included in Appendix A.

In conclusion, there is no way at this time to quantify the area of content validity. It can only be stated that this analysis as presented in table 1 does represent high validity of content for four of the units of study. The fifth unit had low content validity. Consequently, this unit was rewritten to conform with the course content.

Item Analysis of Unit Tests

In order to make each item in each unit test as clear as possible, an analysis of items was undertaken. The available procedures for item analysis of criterion-referenced tests require two administrations of a test. Since this pilot study tested students only once, it was not possible to use those procedures.

A computer program at the California Polytechnic State University scores true-false and multiple-choice. It also prints a percent which indicates the number of students who answered the item correctly. Since there was no other way to identify poor items, the author decided to use the

percent of those who answered the item correctly.

This procedure provided another opportunity to evaluate the test items closely.

Any item answered incorrectly by more than 50% of the students was critically evaluated on the following points: Table 2 shows the results of the item analysis.

1. ambiguity
2. poor grammatical structure
3. more than one answer to an item
4. irrelevancy to the objective

The above points were selected by the author as necessary for clarity of each test item. Each test item was rewritten when it appeared to fail one or more of the points.

The use of the above four-point criteria for item analysis may not fit the philosophy of criterion-referenced testing. Criterion-referenced measures relate an individual's performance to an achievement level which indicates the best performance. The goal of an instructor is to bring each student to a point of optimal performance (mastery of content) without regard for relative group comparisons. Therefore, the type of item analysis just described can be argued as inappropriate for criterion-referenced test procedures. The four-point criteria just described have value to criterion-referenced tests. A low response of correctness for test items in the pilot only alerted the author that there may be something wrong with the items. It was another check on the measurement tool which was used

Table 2 . Item analysis in % correct for each question for all unit tests in Greenhouse Management. *Items of less than 50% correct response were revised.

Question No.	I	II	III	IV	V
1	65%	41%*	35%*	59*	100%
2	88	24*	24*	94	12*
3	94	71	24*	47*	71
4	94	88	71	82	47*
5	88	82	71	100	35*
6	100	71	82	71	65
7	64	100	82	76	53
8	94	47*	82	59	47*
9	82	94	94	82	76
10	94	71	41*	94	12*
11	29*	88	71	82	76
12	41*	76	94		59
13	82	59	88		82
14	18*	41*	65		76
15	88	35*	59		100
16	94	71	82		88
17	100	47*	88		88
18	100	41*	88		76
19	47*	88			94
20	94	47*			88
21	71				71
22	41				65
23	88				94
24	82				35*
25					64

in the research. Items were reviewed and changes were made on the above four point criteria. No data were logged as to what changes were made. The revised questions appeared in the instrument which measured achievement.

The analysis was based on the best judgement of the author. The reason for the analysis can not be defended any more than what was stated. Erickson and Wentling state that other data and personal judgement should play key roles in ultimate decisions about item retention and revision. They do not offer any suggestions on criteria to use.

Reliability of the Achievement Test

Since it was important to know how consistent the various unit tests are, it was important to measure its reliability. Erickson and Wentling (1976) define reliability as the degree to which an instrument (test) provides a trustworthy or consistent measure of whatever it does measure. If an instrument has high reliability it is highly consistent in its measurement. Reliability of a test can be determined by comparing student scores on two administrations of a test. The tests can be the same tests or similar tests. Also, comparison of two halves of a single test can be done. Usually the halves are created by separating the odd and even numbered test items into two groups. The scores of each half are compared. The pilot study consisted of only one administration of one form of each unit test. Therefore, reliability estimates are obtained by comparing two halves (odd and even items) of the test.

It was inappropriate to use the usual formulas for determining reliability coefficients because these formulas are used for norm-referenced tests when variability in scores is desirable. In the case of criterion-referenced testing, the variability of scores is minimized since most students are expected to reach a criterion score. In this pilot, 90% was the criterion. For this reason, it is suggested by Ebel (1973), Hambleton et al. (1978) and Erickson and Wentling (1976) that formulas developed recently for criterion-referenced testing be used. Livingston developed a formula for estimating the reliability of criterion referenced measures. This formula was used to arrive at reliability coefficients for this research. It is written as follows:

$$r_{cc} = \frac{r_{xx}S_x^2 + (\bar{X} - C)^2}{S_x^2 + (\bar{X} - C)^2}$$

Where r_{cc} = criterion referenced reliability

r_{xx} = any one of the classical estimates of reliability

S_x^2 = observed score variance

$\bar{X}$ = observed class mean

C = criterion score set for the class

This formula is an adaptation of the classical formula for estimating reliability (r_{xx}) of a test. The reliability (r_{cc}) of the unit criterion-referenced test is

expressed as a coefficient. A high coefficient indicates that the variance of the correlated test is due to the measure and what it is intending to assess. For example, a reliability coefficient of 0.75 indicates that 56.25% (the square of 0.75) of the common variance is due to the test. Thus, the coefficient assists in answering the question, "Would students obtain similar scores on the same test if the students were to be retested?"

The criterion score (C) of the formula is set by an instructor for a course. In this pilot, 90% was the criterion. The average or mean ($\bar{X}$) was derived from actual scores of an unit test. The score variance (S_x^2) and mean was obtained from a computer program at the California Polytechnic State University. The variance (S_x^2) denoted a measure of dispersion or spread from the mean.

If the criterion score (C) equals the observed class mean ($\bar{X}$), Livingston's formula is the same as classical reliability (r_{xx}). The further the criterion score (C) deviates from the mean ($\bar{X}$), the higher the criterion-referenced reliability. As shown in Table 3, the estimated reliability coefficients vary from 0.6900 to 0.9849. What can be inferred from these estimates? Generally, the coefficient provides a quantitive estimate of the accuracy of the test itself. A coefficient of 0.6900 for unit test IV means that the test questions have 48.53% of their variance in common. The highest coefficient, 0.9849 for

unit test II, has 97% of its variance in common. Borg and Gall (1971) state that coefficients ranging from 0.65 to 0.85 are accurate enough for most test purposes. Coefficients over 0.85 indicate a close relationship between the two variables correlated. We can be confident that a very good relationship exists between variables correlated. If we were to measure a student's level of achievement on future administrations of the same test, we would expect the test to give similar results.

Table 3 . Reliability coefficients calculated by the Livingston formula for criterion-referenced tests. Values are shown for each unit test taken by students during the pilot study.

UNIT TEST	RELIABILITY COEFFICIENT	PER CENT OF COMMOM VARIANCE
I	0.7500	56.25%
II	0.9849	97.00%
III	0.9674	93.58%
IV	0.6967	48.53%
V	0.9050	81.90%

Summary and Conclusion of the Validity and Reliability of the Achievement Test

In conclusion, the individual achievement tests were a useful measurement of the domain of knowledge. According to a subjective review, the content validity of the tests correlated highly with the objectives of the course. This

should be so since criterion-referenced tests should be keyed very closely with stated course objectives. Additionally, a high amount of confidence was placed in the unit tests to provide a consistent assessment of the knowledge, and skills, being measured in future administration of the tests. This was reflected by the medium to high reliability coefficients calculated by the Livingston formula for criterion-referenced test reliability.

Development and Assessment of the Attitude Measure

The researcher constructed the attitude measure during the pilot study. After the measure was fully developed and assessed, it was used as part of methods and materials to complete this research. The attitude measure has been developed by the summated ratings method (Edwards, 1957) described in the next paragraph.

In order to develop the attitude measure, a self-made Likert scale described by Edwards (1957) was used. Two hundred individuals were asked to express their feelings about ornamental horticulture by writing three favorable and three unfavorable statements about ornamental horticulture. One neutral statement was also requested. From this response, 50 favorable and 50 unfavorable statements were selected with the assistance of Edwards' 14-point criteria. (See Appendix B.) The 100 statements were randomly placed on pages with a response set reading: strongly agree, agree, neutral, disagree, and strongly

disagree. The response set received a weighting of 5, 4, 3, 2, 1 so that a statistical analysis could be made. The highest weight was given to the 'strongly agree' term when the attitude question was favorable. Conversely, the 'strongly disagree' term was given a 5 weight when the statement was unfavorable. The 100 statements were scored, and a value of 't' was calculated for all statements. The attitude scores and value of 't' was computed by a computer program located at the California Polytechnic State University. Following Edwards' approximate rule of thumb, a value of 't' equal to or greater than 1.75 indicated a significant statement. Therefore, the author selected as many statements as possible which had the greatest 't' value. The value of 't' is a measure of the extent to which a given statement differentiates between the high-scoring and the low-scoring groups.

As a result of the summated ratings methods of attitude scale construction, it was possible to develop two attitude surveys which most likely gave high response values to favorable and unfavorable statements. Each survey had 22 statements, 11 favorable and 11 unfavorable. All statements were randomly placed on the final survey forms.

The expected value for a strongly positive response to the survey was 110 (22 statements times a weight of 5 for 11 strongly agree and 11 strongly disagree statements). On the other extreme, the lowest score of 22 was obtained

when all statements with a value of 1 were picked . A neutral response is 66.

Since a second administration of the newly developed attitude surveys was not possible, an odd-even split half coefficient of internal consistency was calculated for the two forms of the survey. The attitude surveys are shown in Appendix C. The first form had an estimate of reliability of 0.6298 while the second form is 0.559. In 1966, Barker reported a 0.709 coefficient of correlation for a self-made attitude scale toward school guidance. He considers this value as a preliminary estimate of the alternate form reliability of the scale. In general, Mehrens and Lehmann (1973) state that attitude scales have reliabilities around 0.75. Borg and Gall (1971) indicate the low, medium, and high reliabilities for 18 reported attitude scales to be 0.47, 0.79 and 0.98 respectively. Borg and Gall (1971) state that coefficients around 0.50 (25%) common variance may be a crude estimate of what is being predicted. Based on the available references, the calculated reliability coefficients of the attitude surveys were fair. The reliability accounted for only 31.24% to 39.66% of the common variance. This may result in a considerable restraint on what we find. It could be that if we get no differences in attitude later that it is due to the low reliability. Also, any significant differences among scores must be cautiously evaluated. With a low reliability, a very high

level of significance must be used to show that the magnitude of differences among scores is a true difference. In the absence of any established attitude scale in ornamental horticulture, these attitude scales served as a crude estimate of attitudes.

CHAPTER IV

RESEARCH DESIGN AND PROCEDURES

Introduction

The research was designed so that the following three questions could be answered:

1. Will one criterion produce better student achievement than another?

2. Will one criterion produce more favorable student attitudes toward the course than another?

3. Will one criterion produce more efficient study than another?

In order to partially answer the above questions, several criteria and a control were used as treatment variables. The criteria were 80% and 90% fixed criteria and one criterion called ascending. A fixed criterion remained the same for each unit test. The ascending criterion started at 80% for unit one and increased 5% for each successive unit until 90% was reached. The percent of any criterion was used to calculate the minimum score out of a total score which defined mastery of a unit of study.

The control class was not assigned a required criterion level. Instead, the class was graded on a straight

percentage of total points for each unit. The percentages were 90%, 80%, 70%, 60%, and 50%.

The effect of the treatments was measured by self-made multiple-choice achievement tests on each of five units. The scores of the achievement tests were used to calculate averages for each unit for each treatment group. The averages were used to determine group differences.

The effect of the treatments was also measured by a self-made attitude scale. The score of the attitude scale represented the relative degree of positive attitude toward the course in greenhouse management. The average attitude for each unit for each treatment group was used as a comparison among groups.

Lastly, total study time was used as a measure of the effect of the treatments. The study time was reported by students at the beginning of each unit test. The study time represented all time directly spent on studies in greenhouse management. The average total time on studies for each unit was used to make comparisons among treatment groups.

From the research questions for this study, the following hypotheses were drawn. In each case, the selection of criterion was tested for its effect on achievement, attitude, and total instructional time to learn each unit of study.

1. There will be an interaction between treatments and time for mean achievement.

2. The control group will have the lowest mean achievement score of any group for each unit test over the period of the quarter.

3. The ascending criterion-referenced group, and the 80% and 90% fixed criterion-referenced groups will all receive a higher score on a measure of achievement than students in the control class.

4. The ascending criterion group will receive a higher mean achievement score than a fixed criterion group of 80% or 90%.

5. The 90% fixed criterion group will get a significantly lower score on achievement than the ascending group.

6. The 80% fixed criterion group will get a significantly lower achievement score than the ascending group.

7. There will be interaction between treatments and time for mean attitude.

8. The mastery students under criterion-referenced testing will have a higher mean score on a measure of attitude than students in the control class.

9. The ascending criterion group will have a significantly higher mean score on a measure of attitude than the 80% or 90% fixed criterion groups.

10. Students in the 90% fixed criterion group will have a significantly lower mean score on a measure of attitude than the ascending group.

11. Students in the 80% fixed criterion group will have a significantly lower score on the measure of attitude than the ascending group.

12. The 80% fixed criterion group will have a significantly higher mean score on the measure of attitude than the 90% fixed criterion group.

13. There will be an interaction between treatments and time for the mean time spent on studies.

14. The criterion referenced groups will spend significantly less mean time on instruction than the control group.

15. The ascending criterion group will spend significantly less mean time on instruction than the 80% or 90% fixed criterion groups.

16. The 90% fixed criterion group will spend significantly more mean time on instruction than the ascending group.

17. The 80% fixed criterion group will spend significantly more mean time on instruction than the ascending group.

Experimental Design

This study took the form of an experimental design with multiple treatments. As shown in Table 4, the variable matrix for this study was a two-way design having five

repeated measures. The time or unit of study (see Table 4) was crossed with each treatment group. Since groups received different treatments, the students were nested within a treatment. There was an unequal number of students in class and all students in a class were used in the research. The number of students in the 80% and the 90% fixed criterion groups, 90% criterion fixed group, the ascending criterion group, and the control was 16, 10, 13 and 17 respectively. No students were dropped from the course and all students were used in the research.

Independent Variables

The treatments were the independent variables of the research. They are identified in the variable matrix of Table 4 as the experimental treatment variables.

The treatments were criteria and the control. Criteria were defined by three levels; that is, 80% of total points for each of five units, 90% of total points for each of five units and an ascending criterion. The ascending criterion started at 80% of total points and increased 5% for each unit test until 90% of total points was reached.

A control group was used in the experiment to determine the significance of the treatment levels over the traditional method of teaching as used in this study. This control group was graded on a straight percentage of total points. The percentages were 90% for an 'A', 80% of a 'B', 70% for a 'C', 60% for a 'D' and 50% and below for an 'F'.

Table 4. The variable matrix is shown. The multiple dependent measures are shown for each time for each treatment variable.

		TIME 1 (Unit 1)	TIME 2 (Unit 2)	TIME 3 (Unit 3)	TIME 4 (Unit 4)	TIME 5 (Unit 5)
Treatment Variable	STUDENTS	ACHIEVEMENT SCORE ATTITUDE SCORE TIME ON STUDY	ACHIEVEMENT SCORE ATTITUDE SCORE TIME ON STUDY	ACHIEVEMENT SCORE ATTITUDE SCORE TIME ON STUDY	ACHIEVEMENT SCORE ATTITUDE SCORE TIME ON STUDY	ACHIEVEMENT SCORE ATTITUDE SCORE TIME ON STUDY
Group 1 80% Fixed Low Criterion	S_1 . . . S_i					
Group 2 90% Fixed High Criterion	S_{i+1} . . . S_{2i}					
Group 3 80 to 90% Low to High Criterion	S_{2i+1} . . . S_{3i}					
Group 4 Control	S_{3i+1} . . . S_{4i}					

Control of Internal Validity of Treatments

The four treatment groups, shown in Table 4 did not have random assignment of subjects. Rather, groups were randomly assigned to the above mentioned treatments. Since subjects were not randomly assigned, selection is considered a threat to internal validity. In order to control for this source of invalidity, analysis of covariance was used in the statistics of the research. Hence, age, grade point average,and manner of student selection of the course were considered as covariables. It should be stated here that a class profile was also made to determine the statistical significance of group differences for the covariables mentioned. These results are presented in Chapter V.

Another concern of internal validity arises from the multiple testings of the subjects. This is the effect of taking a test upon the scores of a later test. Since the achievement tests for each unit of study were different, there was not any problem with this threat to internal validity. On the other hand, the attitude scale may be remembered by students. For this reason, two forms of the attitude scale were used to obtain data. With two forms of an attitude scale, each form will not be reused until four weeks have passed. This should minimize an effect of one testing on the other.

There are two factors to consider for external validity. One is the possible artificiality of the

experimental treatment and the students knowledge that they are involved in an experiment. The other is the multiple treatment interference. The former factor had been eliminated by not revealing any knowledge of the experiment to any student. The fact that the students were being tested and graded differently from what they were familiar was explained as the approach used by the particular instructor in the Greenhouse Management course. Repeated measures of the attitude scale and the collection of data on time spent on studies was explained as a tool being used by the instructor for self-evaluation of the course. The latter factor may have some effect on generalization of the experiment. No students in the same class received different treatments. But, the effect of students talking to each other outside of class was considered as a possible threat to generalization. Therefore, there was a chance that student attitudes might have varied because students were discussing the method grading. The researcher did not control for this possibility. There are no other major concerns for the validity of the experiment.

Dependent Variables

The dependent variable of achievement on each unit test was the number of correct responses out of a total possible points. Mastery was achieved when students reached the criterion level assigned to a class. The achievement unit test was administered during a lecture hour following the end of a unit of study. Each unit of study was two

weeks long. A second administration of the test was done when a student failed to receive a score which defined mastery. This was done by arrangement out of class with the instructor.

The attitude score was administered at the end of each unit achievement test. Thus, there were five times when attitudes were measured.

The third dependent variable, total time for study, was reported in writing by students. The data was written onto a standard reporting form. The form was collected at the beginning at each unit test.

An objective test was used to measure achievement on each unit. The number of correct responses for each mastery level treatment group was logged for each unit. Each student score out of the total possible score was used in the analysis. The individual score permits means to be calculated and compared among other treatment groups.

Attitude toward the course was the third dependent variable measured. This variable was assessed by a self-made attitude scale as described in Chapter III. Two forms of the attitude scales were used. This was done to minimize the chance that students might remember how they responded on a previous test. Form A of the attitude scale was administered immediately after the end of unit test one, three and five. Form B was administered immediately after unit two and four. The attitude scale was administered

only after the first try of each unit test.

Time spent on instruction and testing was logged by each student for each unit. The time log was collected as a ticket to take each unit test. In this way, the author was assured of getting the time log. At the time of collection, the log was checked for proper recording of minutes and hours.

The covariables of age, grade point average, and manner of student selection of the course (elective vs. required) will be used to control for initial differences among groups.

Procedures

Population and Sample

The following description is a representation of the type of students used in the study. They are described in great detail so that other researchers could reconstitute a similar group of students and so that generalizations may also be made to a larger population. The population of the research consisted of third and fourth year university students in Ornamental Horticulture at the California Polytechnic State University, San Luis Obispo, California. About 80% of the students were seniors.

The age of the students ranged from 21 to 24 and about 31% were women.

The students were first-time (native) high school graduates and transfer students from community colleges throughout California. Seventy point nine percent of the

students were transfers. This factor was not considered to bias the study since all students had been at the California Polytechnic State University for several years.

Most students selected this course as an elective as compared to a program requirement. Of 56 students in the entire study, 72.72% had elected to take the course.

The prerequisite background did not vary among classes which are involved in the research. All students had taken courses in Fundamentals of Ornamental Horticulture. Table 5 shows the percent distribution of students who have taken the prerequisite courses. The percent of other major courses is also shown. The other major courses are shown so that the experimental groups can be typed precisely. The control group and the ascending experimental group both had about 31% of the class with the pot plant prerequisite while the 80% Fixed and 90% Fixed experimental groups had 18.75% and 20% respectively. Only the 80% Fixed and 90% Fixed showed any background in cut flower production. While there were differences among the experimental groups with regard to the additional course work, the difference did not influence the research. The course information was self-contained and was taught with the Fundamentals of Ornamental Horticulture as the only prerequisite course.

In an effort to further type the population with respect to additional course background, students were asked to check courses taken in accounting, business law survey and other business or management. Table 6 on the

Table 5 . Prerequisite profile of students subjected to the criterion treatments.

	Treatments			
Prerequisite Courses	Control N=17	Ascending N=13	80% Fixed N=16	90% Fixed N=10
Fundamentals	100%	100%	100%	100%
Pot Plant Production	31.25%	31%	18.75%	20%
Cut Flower Production	0	0	12.5%	40%

Table 6 . Additional course work taken by the students in the stated criterion treatments.

	Treatments			
Background Courses Taken	Control N=17	Ascending N=13	80% Fixed N=16	90% Fixed N=10
Business Law	100%	76.92%	93.75%	70%
Accounting I	68.75%	38.46%	62.5%	60%
Accounting II	37.5%	23.07%	37.5%	20%
Other Business/ Management	25%	15.38%	31.25%	60%

next page portrays the courses taken by each treatment group. Most students in any of the treatment groups have taken business law survey before entering Greenhouse Management. All other background courses varied quite a bit among the various experimental groups. In all but the ascending treatment group, the next greater background course completed was Accounting I.

Treatments

A different level of criterion was used in each 50-minute mastery class. The criteria were used in conjunction with a mastery learning strategy. The criteria were 80% fixed, 90% fixed, ascending (80% for the first unit and increasing 5% each unit test until 90% is reached. Additional units were graded at 90% of total points). A fixed criterion was one which had the same standard applied to each of the five unit tests in the quarter.

The textual material of this mastery strategy in Greenhouse Management (Ornamental Horticulture 323) was divided into sections. The sections contained:

1. instruction for completion
2. objectives for each unit
3. a set of review questions and
4. the lectures given by the instructor

Mastery in this research was defined by three elements: instruction, grades and testing. In order to reach mastery or achieve an 'A' on the achievement test for a unit of study, the students had to attain the minimum criterion set for each instructional unit in a given treatment group.

Whenever criterion was not met on the first attempt of any test, the student was provided with alternative instructional assistance for a particular unit of study. The alternatives were, but not limited to, tutors, a different text on the subject, library readings, a restudy of notes or a review of lectures on audiotapes. The second attempt for mastery of a unit of content was considered the final try. There was a total of five unit tests with each test being given at the end of a two week unit of instruction.

The last treatment group or non-mastery group was the control for the research. The non-mastery treatment group received the same objectives as the mastery groups. They were lectured on each unit of study and given the same test questions as the mastery groups.

There were several fundamental differences between the mastery groups and the control group. First, the control group had only one try on each examination. They were not required to reach any particular level of achievement. The earned points on the first try of any unit was the grade for that unit. The score of the achievement test was obtained on a straight percent of total points; that is 90%, 80%, 70%, 60% and 50% of total points for each unit test.

Second, the control group did not receive any benefits of remediation. The results of the unit test were shown to the students at the next class meeting. At that time, students had a brief opportunity to review the test

and check any incorrect answers. But students were not exposed to alternative learning aids to assist in a better understanding of the course content of a unit.

Lastly, student scores on the first and only attempt on each unit test represented the ability of students to more or less learn and retain knowledge in a fixed period of time. This was unlike the criterion groups which were given additional opportunity to understand the course content. Retesting evaluated the improved learning status of students.

Since treatments were applied to an entire class of any treatment, multiple treatment interference was eliminated. Although, it was recognized that future administrations of the achievement test within each class will have an effect on students. This was considered a legitimate carry-over effect since one of the primary hypotheses states that a certain criterion score when applied over a period of time will produce a change in attitude of the student toward the subject. In addition, the time the student needed for instruction and testing was altered as a result of maintaining a certain criterion throughout the course of study. Therefore, the relation of treatment to attitude and time needed for study and learning could not be eliminated.

Instrumentation and Data Collection

The measure of achievement was taken using a instructor-made unit test which covered five-two week sequences in Greenhouse Management. The instructor-made

tests contained a number of multiple choice items that were keyed to the objectives for each unit of study. The test was administered to the entire experimental treatment group during a class period. The students responded to the items by marking its answer on an opscan computer sheet. The sheet had the corresponding letters to choices given on the printed multiple choice test. There was only one answer for each question. The total number of correct responses was the individual's score. There is a minimum number of correct responses needed to reach the particular stated criterion.

There were five teacher-made tests for each of the five units of instruction. The five tests were the same for all the treatment groups. Five additional tests for each unit were available for students who did not reach the specified criterion on the first attempt. The second test was administrated by arrangement outside of class. In the event of a second try, the highest score was accepted as the students score. No further achievement testing was done after the second attempt to attain criterion. The control group receive only one attempt on each test.

The individual scores of students for each unit was entered as data into the computer program. The analyses phase of the computer program compared means for each unit test for each group to test for differences.

While the researcher did collect data on the number of retests taken by each group, this information was not used in this study.

The attitude measure was taken by a self-made attitude scale developed by the summated rating method as described by Edwards. The attitude scale was administered to each treatment group immediately after the achievement test was taken. The students responded to the attitude scale by placing a mark on an opscan computer sheet. The students used a response set of strongly agree, agree, neutral, disagree, strongly disagree to answer each of 22 items on the attitude survey. When the survey was analyzed, the response set received a weighting of 5, 4, 3, 2, 1. When the item on the survey was a favorable item, the highest weight of 5 goes to the strongly agree term. Conversely, when the item was unfavorable, the highest weight of 5 goes to the strongly disagree term. In this way, a person with a strongly positive attitude obtained the highest score of 110 (22 statements multiplied by 5. The individual scores of each student from each treatment group was entered on computer cards and used in the analysis. The average attitude score per unit was used for comparison among treatment groups.

The measure of total time spent on instruction was taken as a student-supplied record of all time spent to study for each course unit. Students received a log sheet for each unit. Upon this sheet was written the category of study (for example, read text, studied notes, tutored with instructor), and the time spent on each category by day. The log sheet was then returned to the instructor at the

time of taking the achievement test. To be assured of receiving this data from each student, the time log sheet acted as a ticket for the test. No time log means no test was taken until the sheet was returned. There were no instances where students failed to bring the time log. The total hours of study time per student from each treatment group was the third datum entered on computer cards and used in the analysis. The average total time per unit was used for comparison among treatment groups.

Data Analysis

The attitude survey was hand calculated by applying the weights of 5, 4, 3, 2, 1 to the appropriate response set for each item on the attitude survey. The weights were totaled and the total for each student tabulated for key-punching onto cards. The time reported by students also was hand totaled for each student for each unit. The total was tabulated for keypunching. The total points correct for each achievement score for each student was keypunched onto cards with the corresponding attitude score and time score. Each keypunched card contained the three dependent variables scores of each student for the five units of study in Greenhouse Management. The keypunched cards were analyzed via the California Polytechnic State University IBM 360. A multivariate analysis of covariance with repeated measures with multiple dependent measures was the general program used. Since there was an unequal number of subjects in each treatment group, a specific program called the Finn Program,

Version IV for multivariance was used in the actual analysis. The alpha level for significant differences among means of dependent variables was set at 0.05.

CHAPTER V

ANALYSIS OF THE RESULTS

Introduction

Data from the achievement scores, the attitude scores and study time were collected for the statistical analysis of the experimental test.

The achievement score for a unit test was the number of responses correct out of the total possible score. The score of each student for each unit test was used in the analysis so that mean achievement scores could be calculated. The mean achievement scores were compared to determine any group differences.

The attitude score for each unit for each student was used so that mean attitude could be computed. The scores were used for a comparison of attitudes for each group.

The total study time for each student for each unit was entered into the analysis so that a mean study time could be computed. The mean study times were compared for any significant group differences.

The grade point average, the required-elective factor, and the age of the student were collected as covariable data to be used in the analysis. The required-elective

factor is defined as the way the student chose the course. One or more of these covariables were thought to influence the outcome of the experimental tests.

Fifty-six students were involved in the study. All students who were enrolled in the classes took part in the experimental test. All the classes were within the same academic quarter. Four distinct and separate classes were used to test the experimental variables and the control.

The experimental variables were the levels of criterion for each class and the control. Two classes had fixed criteria of 80 and 90 percent. In each of these classes, the student had to attain a score equal to the fixed percent of the total points possible. The third class had a level of criterion which ascended. The level started at 80% of total possible points for unit one. The criterion was increased 5% for each successive unit until 90% was reached. The criterion remained at 90% until the end of the quarter.

The control class was graded on a straight percent of total points; that is, 90%, 80%, 70%, 60% and 50% of total points. This was done for each unit test. Since the control class was not required to reach any particular level of achievement, the earned points on the first try of any unit was the grade for that unit.

The experimental conditions were repeated five times at intervals of two weeks apart. The study lasted for the full length of an academic quarter.

The data was run on an International Business Machines 360 at the California Polytechnic State University Computer Center employing the Finn Program, Version IV.

Analysis of Covariates

Grade Point Average

As is shown in the analysis of variance, Table 7 , there were no differences among the grade point averages of any of the experimental groups. The necessary F ratio is so low that statistical comparison of means was not reported. The means are illustrated in Table 8 . So, it was not necessary to use grade point average as a covariate.

Table 7 . Analysis of variance for grade point average.

Source of Variation	Mean Square	D.F.	Observed F	Required F 5%
Total	0.2389	55		
Groups	0.0328	3	.131	2.78
Error	0.2508	52		

Table 8. Means of the four experimental groups on grade point average.

Experimental	Grade Point Average
Control/Non-mastery	2.8488
Ascending Criterion	2.7685
80% Fixed Criterion	2.8763
90% Fixed Criterion	2.8710

Statistical Analysis of the Chi Square Test of the Elective-Required Covariate

Students within this research were classified as elective students or as those who were required to take the course. It was necessary to make a statistical check to determine if the ratio of the elective to required classification was the same for each class. If they were not the same, then it would be necessary to use this data as a covariate.

The Chi Square test of the frequency counts of those who chose the course as an elective and those who must take it indicated that the ratio of the various groups were alike. Tables 9 and 10 illustrate the results.

Table 9 . Summary of data from the four experimental groups based on a 7:3 ratio (elective:required). Ratio was obtained when the pilot study was made.

Source	D.F.	Chi Square
Total	4	0.6475
Pooled	1	0.2755
Heterogeniety	3	0.3720

Table 10 . Summary of data from the four experimental groups based on 41:15 ratio of the observed totals.

Source	D.F.	Chi Square
Total	4	0.3926
Pooled	1	0.0000
Heterogeniety	3	0.3926

The Chi Square for a ratio of 7 to 3 (elective to required) indicated a probability of about 95% that the Chi Square value of this size or larger could come from a homogeneous set of samples from a single population. When tested against an observed ratio of the totals (Table 10), the probability is 99% that the population has a ratio of

41:15, elective to required respectively. Also, the observed ratio shows that there is a 95% probability that the groups were drawn from the same population of students. Therefore, this covariable was not used.

Regression Analysis for the Age Covariate

A regression analysis was used to determine the relationship of age and the dependent variables of student achievement, student attitude and total time it took to study each unit. The regression analysis within the Finn Program showed age not be significantly correlated with any dependent variable. The results of the regression analysis are shown in Table 11 .

Table 11 . Statistics for Regression Analysis for the age covariate.

Dependent Variable	Square Multiple R	Multiple R	F	P Less Than
Achievement	0.0097	0.0987	0.5016	0.4821
Attitude	0.0030	0.0546	0.1527	0.6976
Study Time	0.0153	0.1235	0.7906	0.3782

Since age was not significant, it was not used as a covariate.

Summary of Analysis of Covariates

The analysis of variance for the mean grade point average of the four treatment groups indicated that the group were alike on this measure. Therefore, this covariate was omitted from the analysis.

The chi square test was used to investigate the differences in the proportion of students in each treatment group who chose the course as an elective and those who must take it. All groups had the same proportion of students. Therefore, this covariate was omitted from the analysis.

The age covariate was introduced into the multivariance program. The results of the regression analysis of the program showed that age had a very little correlation to any dependent variable. Therefore, age was meaningless as a covariate.

The following sections describe the multivariate analysis of variance for the dependent variables of achievement, attitude,and study time. First, the analysis was done to determine any interaction of criterion and time. Second, a failure to get interaction permitted investigation of the differences between groups.

The statistical results are presented for achievement, attitude and total study in that order.

Interaction of Criterion by Repeated Measures on Mean Achievement

The overall hypothesis regarding the interaction of criterion by repeated measures on mean achievement was: There will be an interaction between treatments and time for

mean achievement. The results of the multivariance test of interactions indicated an F-ratio of 0.8097 with a probability of 0.6400. The initial decision was to reject the overall hypotheses for interaction.

In addition to this hypothesis, several specific hypotheses were written for the achievement dependent variable. They are:

1. The ascending criterion group will have a progressively higher score on achievement for each unit test over the period of the quarter.

2. The 90% fixed criterion group will have a progressively lower score on achievement for each unit test over the period of the quarter.

3. The 80% fixed criterion will have the next lowest but a moderately stable score on mean achievement for each unit test over the period of the quarter.

4. The control group will have the lowest mean achievement score of any group for each unit test over the period of the quarter.

In order to investigate the above hypotheses, the univariate results were used. The results are presented in Table 12 . According to Cooley and Lohnes (1971) and Finn and Mattsson (1978), one may examine the univariate results for significant interactions when the initial test is not significant. As shown in Table 12 , the results of the univariate analysis of variance did not show significance

at a probability of 0.05. Therefore, there was no interaction of the criterion by repeated measures on mean achievement. As a result of this data, the overall and specific hypotheses noted above for the interaction were rejected.

Table 12 . Univariate results for the criterion by repeated measures interaction on achievement.

Variable	Error Mean Square	Hypothesis Mean Square	Univariate F	P Less Than
Linear	3.2978	8.5954	2.606	0.0615
Quadratic	2.8181	1.8935	0.6710	0.5732
Cubic	3.1516	0.2024	0.0642	0.9786
Quartic	2.5938	2.322	0.8955	0.4498

The results of the main effects were examined next in order to determine any significant difference among treatment groups.

The first hypothesis regarding the achievement score on each unit test was: The criterion treatment groups will receive a higher score on the measure of achievement than the control class.

An analysis of variance was used through the Finn program, Version IV, in order to analyze this hypothesis.

The results of group means and the analysis are shown in Table 13 and Table 14 respectively.

The F statistic for the analysis of variance was significant for all comparisons to the control. Therefore, the hypothesis was accepted. Students taught under the mastery strategy with criterion-referenced testing in this research did score significantly higher in mean achievement for each unit test than students of the traditional classroom approach.

Table 13 . Mean achievement scores for each unit test for the groups under study.

	Control Group	Treatment Groups		
		80%	90%	Ascending
Unit Test	S.D.*	S.D.*	S.D.*	S.D.*
1	21.07(1.62)	22.26(1.31)	22.13(1.37)	21.88(1.40)
2	17.93(2.79)	19.45(1.58)	19.50(2.84)	18.65(3.37)
3	18.38(3.19)	19.79(2.56)	21.25(2.36)	19.88(1.98)
4	19.63(3.93)	21.40(2.50)	23.18(0.96)	22.54(2.69)
5	18.61(2.52)	19.37(2.60)	19.90(1.37)	21.31(1.32)

*S.D. is the abbreviation for standard deviation. All standard deviations are in parantheses.

Table 14 . Univariate ANOVA for the comparison of the mean achievement of the criterion groups to the control.

Experimental Group	Error Mean Square	Hypothesis Mean Square	Univariate F	P Less Than
Ascending	7.01036	127.9246	19.6744	0.0002
80% Fixed	7.03457	133.6831	19.0037	0.0025
90% Fixed	6.70051	72.9300	10.8842	0.0002

The univariate analysis of variance was used to determine differences in the mean score of achievement for each unit test among the criterion groups. The analysis was carried out by comparing each of the fixed criterion groups to the ascending group. The hypotheses for these tests were the following:

1. The ascending criterion will receive a higher mean achievement score than the fixed criterion group of 80% or 90%.

2. The 90% fixed criterion group will get a significantly lower score on achievement than the ascending group.

3. The 80% fixed criterion group will get a significantly lower achievement score than the ascending group.

4. The 80% fixed criterion group will receive a higher mean achievement score than the 90% fixed criterion group.

The results recorded in Table 15 indicate that the ascending criterion is no different than the 80% fixed or 90% fixed criterion on mean achievement of the unit tests. The decision was to reject hypothesis 1.

The individual univariate tests made for hypothesis 1 permitted decisions to be made on the other hypotheses in question. Since the ascending group was no different on mean achievement from the 80% or 90% fixed group, then no further analyses were required for the other hypotheses. Since the previous statistic showed that the groups were alike on mean achievement for each unit test than differences did not exist between the other comparisons. Therefore, hypotheses two, three, and four were rejected.

Table 15 . Univariate ANOVA for the comparison of the mean achievement of the ascending criteria to the fixed criteria.

Treatment	Error Mean Square	Hypothesis Mean Square	Univariate F	P Less Than
Ascending To 80%	7.4849	13.1187	1.7527	0.1967
Ascending To 90%	8.1068	.4457	0.0550	0.8169

Interaction of Criterion by Repeated Measures on Mean Attitude

The overall hypothesis regarding the interaction of criterion by repeated measures on mean attitude score for each unit of study was: There will be an interaction between treatments and time for mean attitude. The mean attitude scores for each group under investigation are summarized in Table 16 .

The F-ratio for the multivariate analysis was 2.9124 for 12 and 129.93 degrees of freedom with a probability of .0014. The test of significant interactions for the multivariate test was significant. The decision was to accept the overall hypothesis.

The significant F value for the criterion by time interaction indicated a different attitudinal response to the course depending on the treatment (level of criterion and the control). There was a change in direction of the attitude as well as a change in magnitude of the attitude score depending upon treatment group. Since the initial interaction was significant, the means of all groups were plotted. Figure 1 illustrates the trend of this plotting. The graphed results were visually examined to analyze trends which occurred over time.

The first specific hypothesis for interaction was: The ascending group will have the most positive mean attitude over the period of the quarter. The ascending criterion group showed an increase in attitude at the close of unit two,but it was not the most positive change. While

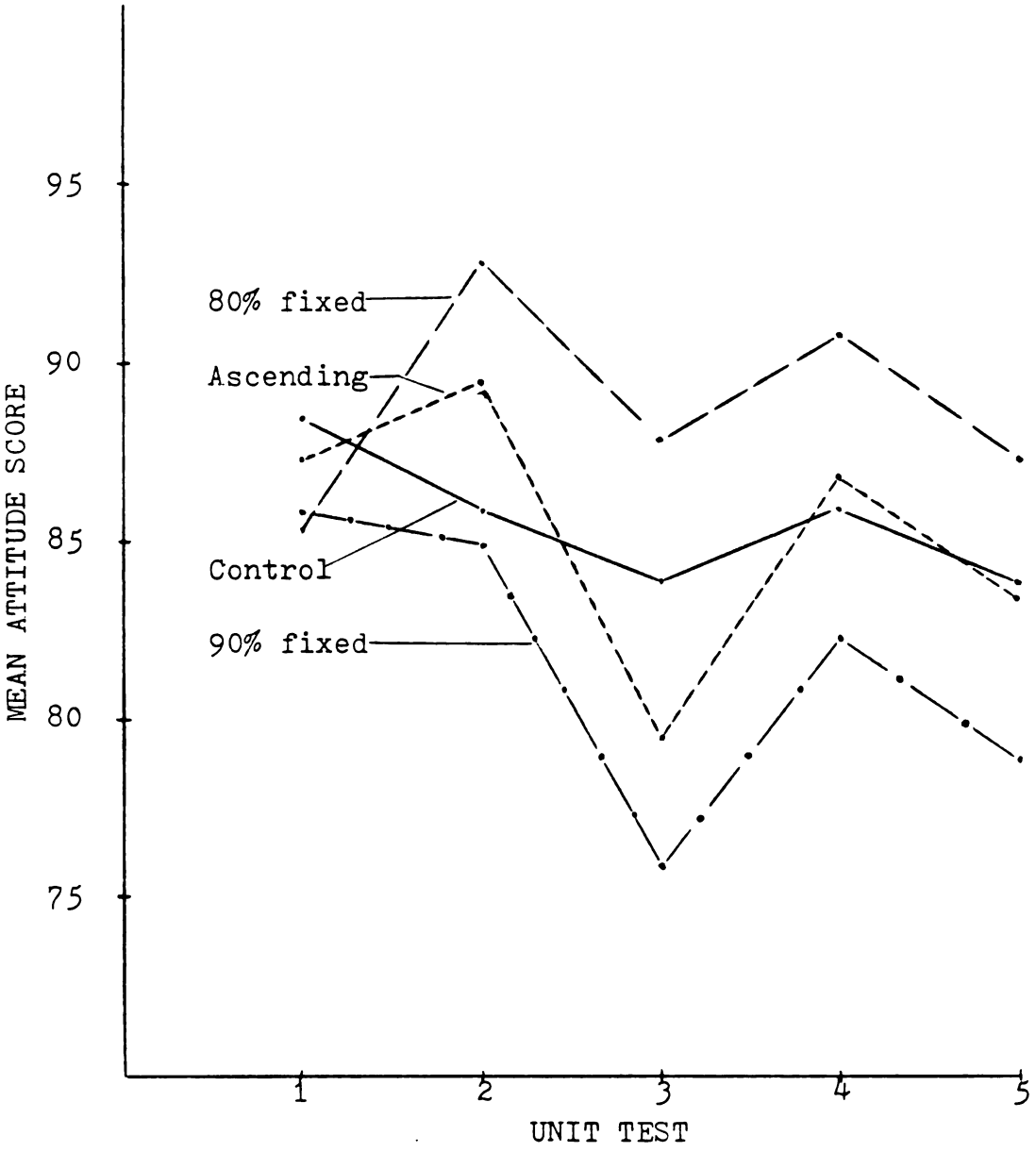

Figure 1. Mean attitude score plotted over time for each group under study.

the attitude of students fluctuated up and down for the remainder of the quarter, the trend in attitude showed an overall decline. Therefore, the hypothesis was rejected.

The second hypothesis for interaction was: The 80% fixed criterion group will have the next most positive attitude over the period of the quarter. The pattern of change of positive attitude increased much faster than the ascending group by unit two. The attitude at this time was the most positive of any group (see Figure 1). After unit two, attitudes fluctuated up and down as the ascending group but not as sharply. Even with these fluctuations, the 80% criterion group continued to show the most positive attitude. Therefore, the hypothesis was rejected.

The next hypothesis for interaction was: The 90% fixed criterion group will have a progressively negative attitude over the period of the quarter. A progressively negative attitude was defined as a general decline in attitude over the quarter. Overall, the trend was for a decreasing student attitude toward the course throughout the quarter. The attitude of students was less negative at unit four, but this correction ended when attitude became more negative again at the end of the quarter. Since the attitude of the 90% group, in general, was progressively more negative over time, the hypothesis was accepted.

The last hypothesis was related to the control group. It was: The control will have the most negative attitude toward the course over the period of the quarter.

The control did not react in the expected direction. Instead, the attitude of the students toward the course was relatively positive and unchanged throughout the quarter. This is easily seen in Figure 1. This was unlike the criterion groups which showed increasing or decreasing positive attitudes depending upon group. The hypothesis for the control group was rejected.

Table 16 . Mean attitude scores for each unit test for the groups under study.

	Control Group	Treatment Groups		
		80%	90%	Ascending
Unit Test	S.D.*	S.D.*	S.D.*	S.D.*
1	88.47(8.40)	85.31(9.88)	85.80(5.98)	87.31(7.33)
2	85.94(6.32)	92.63(9.21)	84.80(4.71)	89.46(6.94)
3	83.94(5.63)	87.63(7.53)	75.90(11.98)	79.46(7.61)
4	85.82(5.35)	90.63(9.02)	82.20(13.52)	86.62(8.06)
5	83.82(7.09)	87.13(8.36)	78.80(10.39)	83.46(7.66)

*S.D. is the abbreviation for standard deviation. All standard deviations are in parantheses.

Since there were significant interactions, the results of the main effects were meaningless. All further hypotheses on mean attitude were ignored.

Interaction of Criterion by Repeated Measures on Mean Study Time

The overall hypothesis regarding the interaction of criterion by repeated measures on mean study time was: There will be an interaction between treatments and time for the mean time spent on studies.

The results of the multivariate test of interactions indicated a F-ratio of 2.0762 for 12 and 129.93 degrees of freedom. The initial test for the interaction was significant at a probability of less than 0.0227.

The significant F value for the groups by time interaction on the measure of time spent on studies indicated a different response to the amount of study time reported depending on treatment level (criterion level and control). Since significance was noted for the overall hypothesis, a specific hypothesis is presented below for each treatment group. The means concerning the treatment group are summarized in Table 17. The means were plotted (Figure 2) so that a visual examination of the trends in study time could be analyzed. Based on this analysis, one could accept or reject the subhypotheses.

The first hypothesis was for the ascending group. It was: The ascending criterion will spend progressively less time on studies over the period of the quarter. Progressively less time is defined as a gradual decline in the amount of study time over the quarter.

As shown in Table 17, the ascending group spent

progressively less mean time in hours on studies throughout most of the quarter. As graphically illustrated in Figure 2, the mean study time did level off somewhat by the end of the 10 week term (unit five). This result supported the hypothesis. Therefore, the hypothesis was accepted.

The second hypothesis was: The 80% fixed criterion will spend a steady amount of time on studies over the period of the quarter. The 80% group did not respond as expected. As shown in Figure 2, the reported total time spent on studies for the 80% group was about the same for unit one and unit two. After this point, total study time declined but the initial rate of decline from unit two to unit three was not as sharp as the ascending group or the 90% group. After unit three, students maintained a low amount of study time. Furthermore, the results indicated that the study time was about the same for unit four and five. Since the results did not support the hypothesis, it was rejected.

The 90% criterion-referenced group was expected to spend more time on studies throughout the quarter. Therefore, the third hypothesis was: The 90% fixed criterion group will spend progressively more time on studies over the period of the quarter. Progressively more time is defined as a gradual incline in the amount of study time over the quarter.

As reported for unit one, the students of the 90% group began the quarter with the greatest amount of time

spent on studies. After that unit, the 90% fixed group declined rapidly in mean study time on subsequent units until unit five. The reported results (Table 17) show that the mean study time increased at unit five. The rate of decline for the quarter was the sharpest of any treatment group. While the 90% fixed criterion reacted in the manner just described, they did not respond according to the expectation of the hypothesis for the 90% fixed criterion. Therefore, the hypothesis was rejected.

Lastly, the control group was examined on the basis of the following hypothesis: The control group will spend the most time on studies over the period of the quarter. The control group had somewhat of an erratic behavior on study time. A decreasing trend in study time was noted in the beginning of the quarter,but the time the students spent on studies rapidly increased at unit three,followed by a large decrease for unit four. After unit four, the time spent on studies increased slightly. In addition, the results were similar to those of the 80% group. These results are shown in Table 17 and graphically presented in Figure 2. Since the general trend of the control group was to spend less time on studies over the quarter, the hypothesis was rejected.

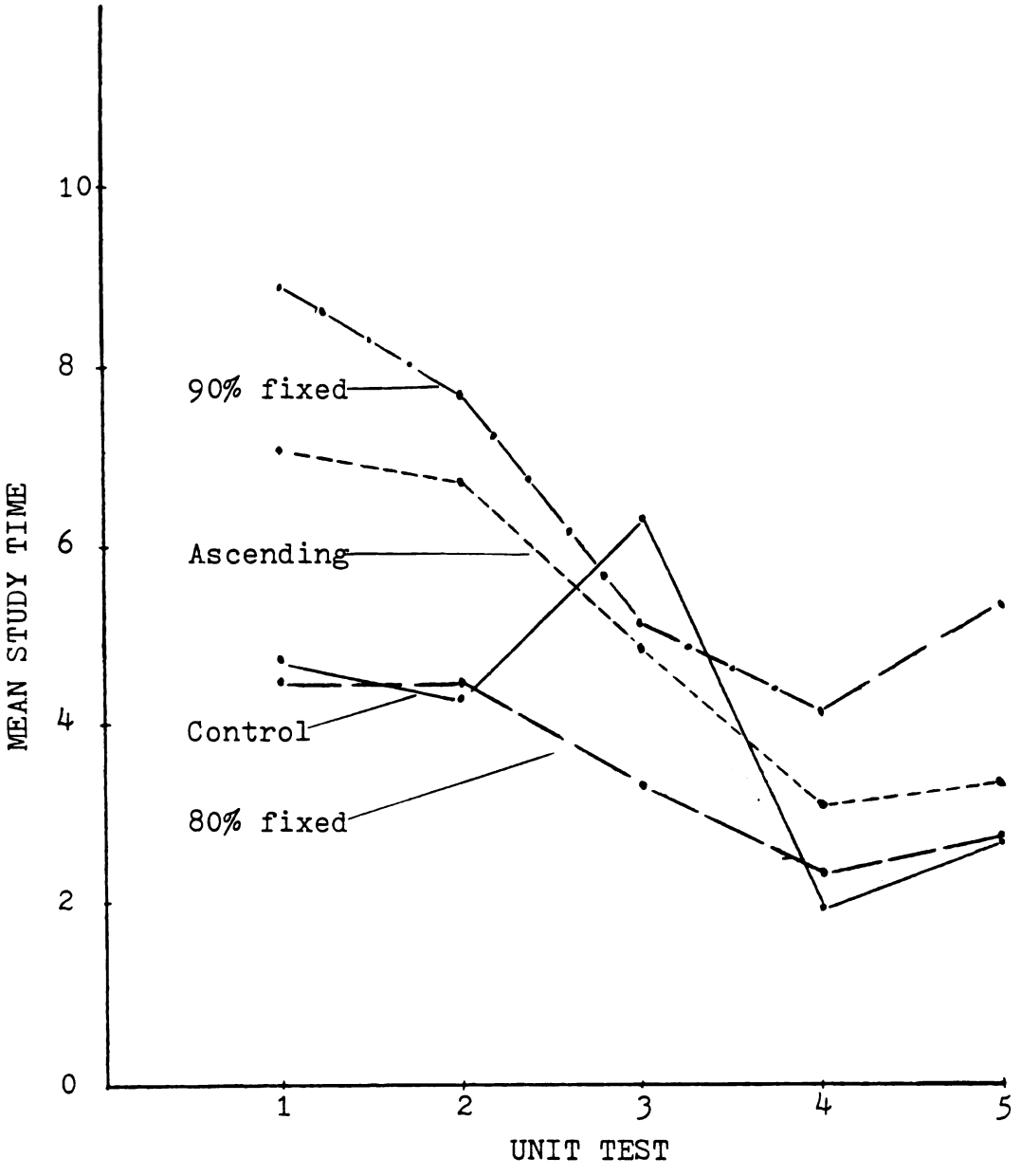

Figure 2. Mean study time plotted over time for each group under study.

Table 17. Mean study time in hours for each unit for the groups under study.

	Control Group	Treatment Groups 80%	90%	Ascending
Unit Test	S.D.*	S.D.*	S.D.*	S.D.*
1	4.69(2.80)	4.49(2.56)	8.96(4.11)	7.05(3.94)
2	4.35(3.26)	4.53(2.33)	7.66(7.43)	6.70(4.05)
3	6.23(4.96)	3.35(1.90)	5.13(4.84)	4.85(3.41)
4	1.96(1.53)	2.31(1.07)	4.15(3.43)	3.12(1.72)
5	2.72(1.97)	2.74(1.35)	5.31(3.23)	3.36(1.64)

*S.D. is the abbreviation for standard deviation. All standard deviations are in parantheses.

Summary

The statistical analysis of the study has been presented in this chapter. Measures of the students achievement, attitude toward the course, and study time were taken for each of five units. The measures were analyzed by a multivariate analysis of variance, Version IV, Finn Program. The computer facilities of the California Polytechnic State University at San Luis Obispo were used during the analyses phase.

There were no significant interactions between criterion and time for the measure of achievement. The achievement scores of the criterion groups were significantly higher than the control throughout the term. There were no differences in mean achievement among any of the criterion groups.

Five hypotheses were written for the interaction of criterion and time on mean attitude toward the course. The overall hypothesis indicated that there was a significant interaction. The following hypotheses were written in order to present the expected direction and magnitude of a response by a particular treatment group within the interaction. The decision for each hypothesis is also presented.

1. The ascending group will have the most positive mean attitude over the period of the quarter. The decision was to reject the hypothesis.

2. The 80% fixed criterion group will have the next most positive attitude over the period of the quarter. The decision was to reject this hypothesis.

3. The 90% fixed criterion group will have a progressively negative attitude over the period of the quarter. The decision was to accept this hypothesis.

4. The control group will have the most negative attitude toward the course over the period of the quarter. The decision was to reject this hypothesis.

Since many of the attitude hypotheses were not supported, the following summary of the actual responses is presented.

The 80% fixed criterion group showed the greatest initial increase in attitude toward the course from unit one to unit two. During the same time, the ascending group showed an increase in positive attitude, but to a lesser extent. The higher positive attitude was not maintained throughout the course by either group. Instead, the attitude was more or less positive until the end of the term. The fluctuating pattern was much greater for the ascending group than the 80% fixed group. While student attitude did fluctuate for the 80% fixed group, the students of this group maintained the most positive attitude throughout the term.

The 90% fixed criterion group had a progressively more negative attitude over the quarter. The pattern was interrupted at unit four. At that time, the student attitude became more positive, but this increase was not continued. Instead, students returned to being negative in their attitude toward the course at the end of the quarter.

The students of the control group were neutral in attitude toward the course. This was maintained throughout the quarter.

Five hypotheses were stated for the interaction of criterion and time on mean time spent on study. The overall hypothesis showed that the interaction of criterion and time

was significant. Four specific hypothesis were written to state the expected direction and magnitude of a response by the particular treatment groups within the interaction. They are shown below with the decision for each hypothesis.

1. The ascending criterion will spend progressively less time on studies over the period of the quarter. The decision was to accept this hypothesis.

2. The 80% fixed criterion will spend an equal amount of time on studies over the period of the quarter. The decision was to reject this hypothesis.

3. The 90% fixed criterion will spend progressively more time on studies over the period of the quarter. The decision was to reject this hypothesis.

4. The control group will spend the most time on studies over the period of the quarter. The decision was to reject this hypothesis.

Many of the above hypothesis were not supported. Therefore, a summary of what did happen is presented below.

Overall, the criterion groups spent a decreasing amount of mean time on studies throughout the quarter. The ascending and the 90% fixed criterion groups showed the greatest decrease in time spent on studies. This was very obvious for the measurements taken at unit two, three and four. For the same period of time, the 80% fixed criterion decreased in mean study time but the decrease was not as rapid. The results of the unit five indicated that the ascending criterion and the 80% fixed criterion had not

changed much in mean study time from unit four. On the other hand, the students of the 90% group showed a marked increase in time invested in studies.

Overall, the students of the control group were spending a lesser amount of time on studies by the end of the quarter. But students did report an unusually large amount of study time for unit three.

Limitations of the Results on Attitude and Study Time

Since the reliability of the attitude scale was low, there are serious questions as to what the results on attitude represent. For example, the fluctuating pattern of attitudes throughout the quarter may have occurred from other factors which confound the results. Therefore, the attitude findings should be viewed with a great deal of caution.

The total time as reported by students may not reflect the relative efficiency of students. If students do poorly on a unit test, they may take more time to study for the next unit test in an effort to succeed. Therefore, there is a possibility that the study time data may reflect an over-reaction to the poor test results of a previous test. This over-reaction may have led to a greater amount of necessary study time.

CHAPTER VI

DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

Introduction

Schools can provide a successful learning experience for most students. The use of criterion-referenced testing under the mastery strategy offers the greatest potential for these students. While this potential is present, it has been hampered by the lack of a sound basis for deciding whether or not a student can be considered a master. A master is a student who has met or exceeded the criterion score set by the instructor. Thus, the student has learned to a sufficient degree. As yet there is no objective manner for setting the level of the criterion which yields the best learning.

If the problem of setting the criterion level remains unsolved, the degree of mastery of many students will continue to be misjudged. The specific amount of skills a student must know cannot be adequately evaluated. Also, the instructor cannot adequately judge how well students have learned.

In an effort to solve this problem, this research was addressed to the kinds of variations of presenting the

criterion to students in order to yield the greatest achievement. Also, the variations of the criterion were presented to students in order to produce the best positive attitude toward a subject. Lastly, the purpose of the research was to identify a criterion which produced effective study scheduling throughout the term. Thus, cognitive learning could be done with a minimum amount of study time.

In order to provide a comprehensive answer to the problem of setting the best criterion, the research sought to investigate the following questions:

1. Does one criterion produce more achievement than another?

2. Does one criterion produce better student attitude than another?

3. Does one criterion produce more efficient study scheduling than another?

The research questions were analyzed by a series of hypotheses. A summary of these hypotheses are presented below. They are:

1. There will be an interaction between treatments and time for mean achievement.

2. The ascending criterion-referenced group, the 80% and 90% fixed criterion-referenced group will receive a higher score on a measure of achievement than students in the control class.

3. The ascending criterion group will receive a higher mean achievement score than a fixed criterion group of 80% or 90%.

4. The 90% fixed criterion group will get a significantly lower score on achievement than the ascending group.

5. The 80% fixed criterion group will get a significantly lower achievement score than the ascending group.

6. There will be an interaction between treatments and time for mean attitude.

7. The mastery students under criterion-referenced testing will have a higher mean score on a measure of attitude than students in the control class.

8. The ascending criterion group will have a significantly higher mean score on a measure of attitude than the 80% or 90% fixed criterion groups.

9. Students in the 90% fixed criterion group will have a significantly lower mean score on a measure of attitude than the ascending group.

10. Students in the 80% fixed criterion group will have a significantly lower score on the measure of attitude than the ascending group.

11. There will be an interaction between treatments and time for the mean time spent on studies.

12. The criterion-referenced groups will spend

significantly less mean time on instruction than the control group.

13. The ascending criterion group will spend significantly less mean time on instruction than the 80% or 90% fixed criterion groups.

14. The 90% fixed criterion group will spend significantly more mean time on instruction than the ascending group.

15. The 80% fixed criterion group will spend significantly more mean time on instruction than the ascending group.

Experimental Design

The experimental design had multiple treatments which were crossed with the five repeated measures. Since the separate classes received different treatments, the students of each class were nested within a treatment. The number of students in the 80% fixed criterion group, the 90% fixed criterion group, the ascending criterion group and the control was 16, 10, 13, 17 respectively.

The treatments were criteria and the control. Criterion was defined by three levels. The levels were 80% and 90% of total points for each of five unit tests and an ascending criterion which started at 80% of total points for unit one and increased by 5% for each successive unit test until 90% was reached. The last treatment was the control. This group was graded on a straight percent of 90%, 80%, 70%,

60% and 50% of total points on the first and only try of each unit test.

There were three dependent variables in the research design. They were the achievement test, the attitude scale and the total study time as reported by the student.

The dependent variable of achievement on each unit objective test was the number of correct responses out of total possible points. The achievement unit test was administered during a lecture hour following the end of each two week unit of study. A second and final administration of a unit test was administered by arrangement outside of class.

The dependent variable of attitude was measured at the end of the first try of each unit achievement test.

The dependent variable of total study time was collected from the students at the beginning of each unit achievement test. Additional study time was collected from students who restudied in preparation for the second and final try of any achievement test.

The Sample of the Research

The sample of the research consisted of 56 third and fourth year university students in Ornamental Horticulture at the California Polytechnic State University, San Luis Obispo, California. About 80% of the students were seniors. The age of the students ranged from 21 to 24 years and about 31% were women.

The students of the sample were first-time high school graduates and transfers from community colleges throughout California. Transfers were 70.9% of the sample.

Most students selected this course as an elective as compared to a program requirement.

Finally, all students had the necessary pre-requisite course of Fundamentals of Ornamental Horticulture.

Four separate classes are chosen for the research. Since individual students could not be randomly assigned to separate classes, the class itself had to be randomly chosen for each experimental group and the one control group. All the students in each class are used in the research.

In order to conduct this research, the mastery strategy was employed. Students are given a complete set of objectives which delineated the content of the material to which the course was addressed. In addition, the course was broken down into segments or units of study. The units covered two weeks of course material before any unit test was given. The unit test was administered in the classroom at the end of each unit. In the event students do not master the evaluative instrument, correctives were offered to the student. These correctives included additional study of notes and/or textbook, further readings in textbooks which are recommended references, listening to

audio-tapes related to the course and tutoring assistance on material which was not understood. The student or students who were classified as non-masters were retested to determine if they have succeeded in the new understanding of the course content of a unit.

Lastly, teaching for mastery demands that a criterion be set which defines whether or not the student can be declared as a master for part of all of the course. One class was assigned a criterion of 80% for each unit test; one class was assigned 90% for each unit test. The last experimental group was assigned a criterion which had an increasing standard. This group started at 80% and rose 5% each unit until 90% is reached. The response by the students to the different criteria offers the opportunity to explain the objective basis for setting the criterion under the mastery strategy as used in this research. The response was measured by an achievement score, an attitude of the student toward the program and a total study time invested in studies.

The control class was the last treatment group. This group received the same objectives as the criterion groups. They were given the same lectures on each unit of study and the same test questions as the criterion groups. They differed in several ways. First, they were not required to achieve a level of criterion which defined mastery.

Second, students were not given remediation for their learning inadequacies. Third, retesting was not done to re-evaluate student performance.

Method of Data Collection

The measure of achievement was taken using an instructor-made multiple-choice unit test which covered five two-week sequences in Greenhouse Management. The test questions were keyed to the objectives for each unit of study. The total number of correct responses was the individual's score. The analysis phase of the computer program compared means for each unit test for each group in order to determine group differences on achievement.

The attitude measure was taken by a self-made attitude scale developed by the summated rating method as described by Edwards (1957). The attitude scale was administered to each treatment group immediately after the achievement test was taken. A student with a strongly positive attitude could obtain a score of 110. A strongly negative attitude was measured at 22. A neutral attitude was measured at 66. The average attitude score per unit was used for comparison among treatment groups.

The measure of total time spent on studies was taken as a student-supplied record of all time spent to study for each unit. When additional study was required as a result of not reaching mastery on the first test of any unit, the additional time for that study was also

reported. This time was added to the rest of the time each student spent on studies for each unit. The average total time per unit was used for comparison among treatment groups.

The Importance of the Covariables

The covariables: age of the student, method of course selection and grade point average were selected as possible factors which could bias the data. It was thought that these variables may influence the results of the research. Therefore, the treatments may not be the sole variable influencing the results. In order to avoid this problem, the covariables can be removed statistically so that the effects of the treatments can be analyzed.

The results indicated that students were no different on age, grade point average and method of course selection. The results may be generalized to students of an age group of 21 to 24 years. Furthermore, the results may be generalized to a group of students whose range of average grade points was 2.76 to 2.87 out of a 4.00. Thus, the entering achievement level of the students in each group was the same.

Lastly, all the groups had 70 per cent of the students selecting the course as an elective. Therefore, a sample with a ratio of 7 to 3 will be likened to this sample. If the ratio would have favored a required selection of the greenhouse management course, then one should expect those students to respond differently to the attitude survey.

It seems reasonable to suggest that students who must take a course have a different attitude toward the course than students who want to take it. Furthermore, there is a chance that students who are required to take a course will be less motivated to achieve to a level that they are capable. These students are often classified as underachievers. It also follows that underachievers probably spend less time on their studies. With this in mind, the study time data of this research may be confounded.

In summary, the samples of students in each of the experimental groups involved in this research are homogeneous with regard to age, grade point average and method of course selection. As a result, one can expect to draw students from such a population and get similar results.

Discussion of the Analyses of the Results

Discussion on the Results of Achievement

The use of any criterion as used in the mastery strategy of this research did produce higher achievement scores than the control group. These findings should be viewed in relation to the similarities and differences of the criterion groups and the control group.

The criterion groups and the control were similar with regard to the following procedures. The groups were given instructions for completion of the course, objectives for each unit and a set of review questions for each unit of study. All instruction was done in class to an entire group.

The differences between criterion groups under the mastery strategy and the control may have produced the higher achievement for the criterion groups. First, the mastery strategy of this research required students to reach a particular level of achievement. For two of the three mastery groups, the levels of achievement were 80% and 90% of total points for each of five unit test. The third group had an ascending criterion which started at 80% and increased 5% for each successive unit test. Therefore, attainment of an achievement score, which defined the criterion, would indicate that the best learning of course content has taken place.

Second, the significantly higher achievement scores for the criterion groups may also be attributed to the benefits of remediation which was employed in the mastery strategy of the research. If students failed to reach a certain criterion level, then certain alternative learning resources were prescribed to students so that a better understanding of the material might occur. These alternative learning resources were additional reading materials on a particular course objective or objectives, a review of audio-tapes for a particular segment of a unit, tutoring assistance or a review of the class notes of a student.

Lastly, the strategy of mastery permitted further assessment of the adequacy of students' learning on a unit of study in which there was a deficiency. Therefore,

retesting was done to evaluate the improved learning status of a student.

A comparison among the criterion groups without the control sought to answer the following question: Does one criterion produce more achievement of the course content than another? The results of this research have shown that none of the treatments using variations of criteria were any different from each other on achievement. Accordingly, all students were alike in their learning of course content.

One may interpret these results as meaning that the mastery approach of this research was able to bring students to a minimum level of performance of the 80% fixed criterion. A visual examination of the means in Table 13 shows that the averages were similar to the group subjected to an 80% criterion. Setting a criterion any higher than 80% did not produce higher achievement. Therefore, it may not be necessary to set a higher criterion under the circumstances and subject matter described.

In summary, criterion-referenced testing under the mastery strategy produced higher achievement than the control group. The reason for the higher achievement may be credited to learning under the mastery strategy. The strategy is designed to bring most students to a better understanding of the course content. In order to meet this goal, students were presented with the results of their test so that deficient areas could be noted. After this,

students were given the necessary alternative learning resources which permitted further study of objectives not understood. After additional studying was completed, students were retested to evaluate their overall understanding of the objectives of a unit of study.

A comparison of the criterion-referenced groups showed that the groups were alike on achievement of unit tests. An examination of the means (Table 13) for all criterion groups has revealed that the group means were like the 80% group. Setting higher levels of criterion would appear to be unrealistic under the mastery strategy conditions of this research.

Discussion on the Results of Attitude

Attitude of the student was introduced into this research as a dependent variable because it was thought that certain criteria may influence the attitude of the students toward the course. Therefore, it was necessary to answer the question: Does one criterion produce better student attitude than another? The results have shown that there was a significant interaction between criterion and the repeated measures on student attitude toward the course in greenhouse management.

When the relatively unchanged attitude of the students of the control was compared with the ascending criterion and the 90% fixed criterion, it appears that the overall attitude of the students of the ascending group and

the 90% fixed group were becoming more negative over time. Furthermore, the attitude fluctuated up and down throughout the quarter. One should have expected that students of the ascending criterion and the 90% fixed criterion who were given the opportunity to do better in their course work should have had a better attitude about the course in which they were enrolled. This was not the case in this research. The setting of a high criterion either from the beginning of the course or gradually working up to it as in the case of the ascending criterion may have produced much frustration in an attempt to achieve at such a high level. This idea was further supported by the lack of significant differences among the achievement of the scores of any of the criterion groups. Even though the students were alike on achievement, the students who were pushed to attain high achievement scores reacted by becoming more negative toward the course. It may be that students though that the higher criterion was an unreasonable expectation of academic success in greenhouse management. Therefore, they became more negative as the term progressed.

The 80% fixed group showed fluctuations in its attitude toward the course. While attitude fluctuated, it was the most positive attitude of any group. The relative ease of attaining criterion for each unit may have produced a high positive attitude toward the course. Additionally, the students were probably not threatened with the prospect

of achieving to an unreasonably high level. Their high academic standing was very secure and therefore there may not have been any reason for students to get a negative attitude.

The fluctuating pattern of attitude for all the variations of criterion may have resulted from a difference in the difficulty of one test or another. The tests were constructed with the intention of one not being more difficult than another, but there was no measurement of difficulty. One might argue that if the difficulty of a test influenced the results, that attitude score could have been influenced by the relative ease by which criterion groups could reach a level of performance. This could account for the degree of fluctuation of attitude throughout the quarter by the criterion groups. When a criterion was easier to reach, the attitude tended not to change as much. This argument is weak because the control group did not react to any apparent difference in test difficulty. Therefore, the author suggests that the results were more likely due to the level of criterion and the requirement that students reach it.

Alternatively, various personal circumstances of students on the day of the test might have changed their attitude. But this argument is even weaker than the above because all students in the different criterion groups could not have had simultaneous good or bad days.

Furthermore, the control group would have reacted in a similar fashion. But the control did not show fluctuations in attitude as did the criterion groups.

The results on attitude were further complicated by the low reliability of the attitude measure. There was a chance that the variations in attitude or the lack of variation was due to an inconsistency of what the scale intended to measure. For example, the unchanged attitude of the control group could have happened from a lack of accuracy of the scale to detect attitudinal change over time. On the other hand, the fluctuations of attitude in the criterion groups may have resulted from the scale measuring something other than attitude toward the course. For example, students might have felt that the method of instruction under which they were taught was unreasonable.

In summary, the overall trend in student attitude toward the course varied according to the treatment group. By setting a low criterion of 80% the attitude of students was the most positive of any treatment group. The 80% fixed criterion produced the best achievement in the course without sacrificing the positive attitude of the students toward the course. It may be that the relative ease at which students achieved on unit tests produced a positive attitude toward the course.

The most negative attitude toward the course over time was noted for the 90% fixed criterion group. Also, the general trend for the ascending attitude was to be

progressively negative in attitude over time but not as much as the 90% fixed criterion. It may be that when students were pushed to attain high levels of achievement from the beginning of the quarter or gradually working up to high levels of achievement, the effect was to produce a negative student attitude toward the course in greenhouse management. Also, if students felt threatened by a criterion which was difficult or impossible to attain, the response may have been a negative perception of the course.

Discussion on the Results of Total Study Time

Since students of the mastery strategy were subjected to additional study and retesting when their performance on a unit test did not reach a level of criterion set for a treatment, a dependent measure of total time on studies was made. This total study time is a measure of relative efficiency among the treatment groups. Thus, the research sought to answer the following question: Does one criterion produce more efficient study scheduling than another?

A significant criterion by time interaction on the dependent measure of study time indicated that there were differences in the amount of total study time among the treatment groups through the quarter. The differing amounts of time accounted for the change in direction of the plotted total study time as shown in Figure 2 .

The trend was for an increasing efficiency of

studies which was noted by a general decline in study time for each successive unit. While efficiency increased over time, the amount of time spent on studies varied with treatment groups.

As the results generally have shown, the 90% fixed criterion and the ascending criterion declined in study time over the quarter. But the time spent on studies was still more than the 80% criterion or the control. The students of 90% group spent the greatest amount of time over the quarter. This was followed by the students of the ascending group. The decline in total study time did not persist for the 90% fixed group. Instead, there was an increase in study time at the end of term.

The trend by the 90% fixed criterion group is explained in this manner. The students of the 90% group were faced with a very high level of criterion. Because of this high level, the students apparently put more time into studies than any other treatment group in an attempt to reach the criterion set for each unit test. But a higher level of achievement was not reached as a result of adding additional hours of study as compared with the other criterion groups. There were no significant differences on achievement among criterion groups as has already been stated earlier.

The increase in total study time at the end of the term might have resulted from students being insecure about

doing well on the last examination of the quarter. Since this group may not have been sure about their ability to reach the criterion score at the end of term, more hours were used to study. Thus, while there was a trend to use less total time to study by the students of the 90% fixed group, the study efficiency was not maintained at the end of the term.

The ascending criterion and the 80% fixed criterion were alike on the level of criterion set in the beginning of the term. However, the ascending criterion spent more total time on their studies at that time. But the total study time was less than the 90% fixed group (see Figure 2). While the level of criterion was easier to attain at that time, the students of the ascending criterion may have felt that it was necessary to obtain the highest achievement score possible early in the term. Faced with the uncertainty of doing well later in the quarter when the criterion would be higher, the students may have studied alot more in order to do their best on the earlier unit tests. Therefore, additional hours of study were spent in a hope of learning as thoroughly as possible.

When the level of criterion for the ascending group was increased to 90% at unit three, the ascending criterion group became similar to the 90% fixed group on total study time. Subsequent to unit three, the ascending group showed further decreases in total study time and finally a leveling

off at the end of term. On the other hand, the 90% fixed group was tending to increase in total study time during the latter period of the quarter. Since the ascending group was able to spend less time than the 90% group when the level of criterion was kept at 90% then the ascending group probably became more efficient later in the quarter. These results can be interpreted in the following way. When the level of criterion for the students of the ascending group was lower, there was an opportunity to become adjusted to studying for the course. Once their study routine was established, the students did not have to put in any more time than necessary in an attempt to reach the level of criterion set for the course. This is graphically illustrated for these groups in Figure 2. Thus, the study time levelled out at a low point and stayed that way until the end of term. It appears that gradually ascending the criterion has some benefit in producing more efficient study.

Except for the unusually great amount of time spent on studies at unit three by the control, the 80% criterion group and the control were similar on the decreasing time spent on their studies throughout the quarter. The 80% group had a significantly higher achievement than the control with about the same amount of total study time over the quarter. Thus, the students of the 80% criterion were more efficient in total time spent on studies.

Also, the lack of significance among the scores of the criterion groups suggested that the 80% criterion is the best criterion to set in order to produce the best achievement in a time-saving manner.

In summary, there was a general trend for a decreasing amount of total time spent on studies over the quarter. While the time decreased as the quarter progressed, the amount of time spent on studies varied with the treatment group.

The 80% fixed criterion group had the least amount of time spent over the quarter. Also, the 80% fixed criterion group was generally like the control on total study time over the quarter. Since the 80% group achieved more than the control class over the quarter with the same amount of study time, the students were more efficient in their study scheduling.

Also, the reported results of no difference among the achievement scores of the criterion groups suggests that the 80% group was the most efficient of any criterion group in their total study time. This is graphically shown in Figure 2 .

On the other hand, the 90% group spent the greatest amount of time on studies even though it tended to decline over time. The decline in total study time did not last until the end of term. After unit three, there was a general trend to increase the time spent on studies in order

to achieve as much as the other criterion groups. The students studied a great deal for each unit test in an attempt to reach the high level of criterion set for the course. Even though a great deal of time went into studying for each unit, the students did not achieve any more than the other criterion groups.

The ascending criterion group began the quarter with the second greatest amount of time spent on studies. It was at that time that the level of criterion (80% for unit one) was the same as the 80% fixed criterion. Even though the criterion was the same, the ascending group apparently spent more time in an effort to achieve a high achievement score. Since the criterion of the ascending would become higher later, the students might have felt compelled to get the highest possible score early in the quarter.

As time passed in the quarter the total time students spent on studies became less and less. Later in the quarter, total study time became more like the 80% fixed criterion group. The latter trend suggests that the students had a chance to establish effective study scheduling early in the term. Later in the quarter, it was not necessary to use any more time than necessary to attempt to score well. If this was not the case than the ascending criterion should have reacted more like the 90% fixed criterion later in the quarter. This group had a trend

upward in total study time toward the end of the quarter.

CONCLUSIONS

The findings of this research have led to the following conclusions. The conclusions are presented for the dependent measures of achievement, attitude toward the course and total time spent on studies in respective order.

1. The conditions of criterion-referenced testing under the mastery strategy of this research produced significantly higher achievement when compared with the control. Those conditions included the following:

a. The course was divided into units of study.

b. Each unit of study had specified objectives. These objectives stated what the student is expected to learn.

c. Testing was done to evaluate the knowledge learned for each unit.

d. The student was evaluated relative to the performance on a unit test. Therefore, a level of criterion defined mastery or the adequacy of learning on a unit.

e. Students were given an opportunity to review their test and note a deficiency in learning.

f. Those who fail to master the content were given alternative learning resources. In this way further study could be done on the content of a unit.

g. Retesting was done to determine if mastery of content has been reached.

2. Under the mastery conditions of this research, achievement was the same regardless of the level of criterion. These findings have shown that higher achievement was not produced by setting a higher criterion throughout the term. In addition, a gradual increase from 80% to 90% (ascending criterion) did not yield any more learning than the other groups. Based on these findings, it appears that the level of criterion to set for instruction like that in this study should be 80%. Above that point, students were unable to master any more material.

3. Because the control group had a relatively unchanged attitude over time as compared with the decreasing positive attitude of the ascending criterion and the 90% fixed criterion groups, it can be concluded that students who were pushed to attain high levels of criterion may become negative in their attitude toward the course as time progresses. This trend was not shown by the 80% fixed criterion group. The higher positive attitude of the students of the 80% group throughout term suggest that it may be better to set the criterion lower to produce the best achievement and the most positive attitude toward the course. The lack of significant difference among achievement scores of the criterion groups suggests that the push to attain higher levels may have produced a decline in positive student attitude toward the course.

4. The setting of a high fixed criterion (such as 90%) forced students to spend a great deal of time over the quarter in an attempt to reach the high level set for the course. This was of no avail since students were not able to reach the high level of achievement set for the course. The criterion may be an unrealistic standard to set under the conditions of mastery as used in this research.

5. There is no reason to believe that gradually raising the criterion under mastery learning has any advantage in producing more achievement in a time-saving manner than a lower fixed criterion. The ascending group eventually decreased in average total study time to approximately the averages of the 80% fixed group in order to achieve as well as the 80% group. Thus, it may be better to set a relatively low criterion (such as 80%). This produced the best efficiency of time spent on studies throughout a quarter.

6. In general, the similar trends and total study averages between the 80% and the control suggests that setting an 80% fixed criterion strategy may be more efficient in producing higher achievement under mastery learning than the control or non-mastery group.

Summary of Conclusions

In summary, the conditions of criterion-referenced testing under the mastery strategy produced significantly higher achievement when compared with the control.

The research also indicated that setting a higher criterion or gradually raising the criterion did not yield higher achievement than a lower criterion. The student attitude toward the course was also less positive when they were pushed to attain higher levels of criterion. Also, higher fixed criterion (90%) produced less efficient study scheduling than other criteria without any gain in achievement over a lower criterion. While the trend in the ascending criterion (start at 80% and increase 5% to 90% and then hold at 90%) was for increasing efficiency, there was not any commensurate gain in achievement over the other criterion groups.

The findings of this research suggested that the 80% fixed criterion could produce the best learning without sacrificing the attitude of students toward the course. In addition, a criterion which was set at 80% yielded the best student learning in the least amount of total study time over a quarter.

RECOMMENDATIONS AND FURTHER QUESTIONS

The following recommendations and questions for further research are based upon experiences of this research project. They are:

1. The setting of a relatively low criterion of 80% produced the best achievement. However, there may have been an adverse effect of the low criterion. The students may not have learned enough material in the quarter. To

overcome the problem of the learning of an inadequate amount of material, a different criterion may be necessary to assure that an adequate amount of material is learned. Would a criterion of 85% yield the best learning?

2. The degree of difficulty of a particular criterion may be expressed as the number of students who master units of study. This type of data may be useful for setting an optimum criterion. Therefore, the following question may be answered: Is there a criterion which produces the greatest number of masters?

3. Data on the number of retests taken by students at the various levels of criterion should be taken. This may suggest the relative difficulty of a criterion. In this way, the following question may be answered: Is there a criteria which produces the lowest number of retests among students?

4. The attitude scale in future research should be revised to include a subset on attitudes toward the method of instruction under which the students are being taught. In this way it may be possible to measure the attitude of students toward the use of criteria under the mastery strategy. Would the use of criteria produce attitude changes in students who are taught by the mastery strategy?

5. An attitude scale with a low reliability may produce results which are not a consistent measure of

student attitudes toward the course. An attitude scale with a minimal reliability of 0.75 should be used to assure accuracy in measurement of student attitude. With higher reliability, one would have a great confidence in the predictability of the results. Consequently, the following question may be answered with greater accuracy: Is there a level of criterion which produces the best attitude toward a course taught under the mastery strategy?

APPENDICES

APPENDIX A

Course Objectives

Greenhouse Management (one quarter)

Define management

1. Given the term management, the learner will define management as developed from the class discussion.
2. Given the term management, the learner will define the term as it is correctly identified to line management, maintenance, production management and sales management as stated in the class.
3. Given the term manager, the learner will define the term as it was developed in class.
4. Given statements or components of statements for management or manager, the student will identify the statement as it applies to management or manager.

Describe the types of business ownership

1. Given the terms sole proprietorship, partnership and corporation and without reference material, the learner will describe them in writing as to include the considerations, liabilities and limitations of the terms with 100% accuracy.

2. Given the terms general partnership and limited partnership and without reference material, the learner will describe as to liabilities and limitations of the terms with 100% accuracy.
3. Given the article terms for corporations and partnerships, the learner will define in writing that term with 100% accuracy and as presented in class.

Describe different ornamental horticulture marketing set-ups

1. Given the terms wholesale house, pool system, cooperative and direct sales on a work sheet, the learner will describe in writing each term as it is applied to the marketing of ornamental horticulture crops according to lecture.
2. Given the terms wholesale house, pool system, cooperative and direct sale, the learner will list secondary advantages to the ornamental horticulturist in columnar form for each term according to lecture.
3. Given a statement or phrase which best describes the marketing set-up, the learner will choose the best term from a group which identifies the statement. This will be with 100% accuracy.
4. Given a key phrase as marketing channels, the learner will diagram the channels for the sale of ornamental horticulture crops according to lecture and references used to develop lectures.

Apply marketing set-ups to ornamental horticulture firms

1. Given the marketing set-up terms and the ornamental horticulture firms, the learner will identify each firm with the appropriate marketing set-up term with 100% accuracy.
2. Given the marketing term 'auction selling' and without reference material, the learner will apply this marketing procedure to the sale of cut flowers and potted plants in the U.S. as presented in class.

Diagram organization flow charts for various ornamental horticulture business types

1. Without reference material, the learner will diagram the three types of business organization in a hierarchical manner labeling each level within the diagram with 100% accuracy.
2. With the aid of the above diagram, the learner will diagram with arrows the manner in which communication and responsibility flows with 100% accuracy.
3. Given the concept of communication within business, the learner will explain in writing how it is best achieved in business. This will be done according to class discussion and the text on Greenhouse Management.
4. Given the concept of responsibility within business, the learner will explain in writing:
   a. its role in the organizational charts
   b. its relationship to communication

c. its relationship to business activities within the organization

This will be done according to class discussion and the text on Greenhouse Management.

END OF UNIT I

Identify different recruitment procedures

1. Without reference material, the learner will identify procedures to follow in order to recruit the following employee types:
   a. unskilled worker
   b. foreman
   c. assistant manager
   d. truck driver
   e. production manager
   f. agricultural economist

Describe the orientation procedures

1. Without references, the learner will write reasons for orientating workers according to lecture notes.
2. Without references, the learner will write items which represent areas of orientation according to lecture notes.
3. Without references, the learner will state in writing the relationship between orientation and business efficiency according to lecture notes.

Plan training procedures for ornamental horticulture workers

1. With references and job forms, the learner will prepare

a list of steps for getting ready to train employees including shortcut steps for the employee for an actual job according to the procedures of the Agriculture Education Department of the University of California at Davis.

2. Given references, job forms, greenhouse facilities, tools, soil, plants, etc., the learner will plan an actual training session with a member of the class as the trainee. The training session will be judged on a scale of 1 to 5 for each of the following:
   a. effectiveness of training
   b. training under actual conditions
   c. clarity of training
   d. preparing the worker
   e. preparing the job
   f. the steps listed on the one-page leaflets given in class
3. Given an on-the-job problem situation, the learner will identify the problem as it relates to orientation or training. This will be done according to handouts and lecture notes.

END OF UNIT II

Estimate of production peaks for ornamental horticulture crops

1. Without references, the learner will calculate the expected date of peak production by random counting of

selected shoots or flower buds or stages of growth with 90% accuracy. (Accuracy is based on previous crop records.)

2. With the necessary references, the learner will graph or write the dates of estimated peak production for a one-year period with 100% accuracy.
3. Given references, the learner will estimate the date of flowering of Easter Lilies according to the standard leaf and bud counting procedures in print.
4. Without references, the learner will state in writing the stages of plant growth which decide when crops will peak in accordance with the stages shown in laboratory on living plants.

Calculate year around crop rotations for selected crops

1. With references and crop rotation forms, the learner will calculate year around pot plant rotations for:
   a. the appropriate number of crops
   b. utilizing 365 days of the year
   c. for specific holiday periods
   d. utilizing 85% of the space each month

   All forms to be returned in an appropriate, specified period of time.
2. Given references, a list of cut flower crops or nursery crops, holiday names and dates, the learner will calculate the correct planting dates, pinching dates and bloom dates based on the first expected holiday bloom within a two-day accuracy.

Calculate number of plants/pots

1. Without references, the learner will calculate the number of pot needs in area to the exact value as computed by the prescribed formula.
2. Without references, the learner will calculate the number of pots in the width and the number of rows in the length. This is in accordance with Weiss formula.
3. Without references, the learner will calculate the number of plants in a bed according to formulas presented in class with 100% accuracy.

END OF UNIT III

Schedule year around ornamental horticulture crops

1. Given reference sources, the student will write a schedule for different nursery crops. The schedule is correct when the crops are available at the same time.
2. Given references or cues, the student will schedule in writing a floral crop so that the crop is available:
   a. on the specific holiday or peak date
   b. on a weekly, biweekly or monthly schedule
   c. during a season
3. Without references, tight spacing will be calculated when it is necessary to get maximum utilization of space.
4. With the use of assigned references, the student will write a chart showing:
   a. start and stop dates for each crop
   b. total time of production in days, weeks or months

c. the holiday or season for each crop

The chart is correct when it conforms with all references and commercial practices.

END OF UNIT IV

Compile cost estimates for greenhouse construction

1. Given references, forms and greenhouse blueprints, the learner will compile a cost estimate for all necessary costs of construction material by using a local lumber company as an estimator with 90% accuracy.
2. Given references, forms and the names of greenhouse owners, the learner will compile a cost estimate for heating systems, cooling systems, all plumbing, all electricity and growing tables with 90% accuracy.

Analyze selected total cost of production

1. Given references, the learner will write a comparison of selected cost of production for different regions of the country. This will be correct when a chart shows the items with the cost for each region.
2. Given references, the learner will state in writing the ratios or % which indicate the financial condition of the company. This is correct when it is in accordance with the Operating Cost Studies of the Horticultural Research Institute.

Define financial term

1. Without references, the learner will write definitions for terms in accordance with the operating cost studies

of the Horticultural Research Institute.

Compute cost of production for selected crops

1. Given a list of crops, cost of production materials, reporting forms and previous assignments, the learner will compute the cost of production by using the reporting form for each crop until all information for each crop is recorded as requested on the reporting forms. The forms are to be returned in an appropriate, specified period of time. All cost figures must be within three decimal place accuracy.
2. Given the computed cost of production for selected crops, the learner will compute:
   a. the selling price for each crop in order to break even
   b. the total cost per sq. ft. per crop
   c. the total cost per sq. ft. of production area per year
   d. the gross receipts of all crops per sq. ft. per year
   e. the net receipts of all crops per sq. ft. per year to the nearest .01 dollars

Describe profitability of crops

1. Without references, the learner will write ways to manage profitability for ornamental horticulture crops according to Perkins and Levins.

END OF UNIT V

APPENDIX B

Edward's Criteria for Selecting Attitude Statements

1. Avoid statements that refer to the past rather than to the present.
2. Avoid statements that are factual or capable of being interpreted as factual.
3. Avoid statements that may be interpreted in more than one way.
4. Avoid statements that are irrelevant to the psychological object under consideration.
5. Avoid statements that are likely to be endorse by almost everyone or by almost no one.
6. Select statements that are believed to cover the entire range of the affective scale of interest.
7. Keep the language of the statements simple, clear and direct.
8. Statements should be short, rarely exceeding 20 words.
9. Each statement should contain only one complete thought.
10. Statements containing universals such as all, always, none and never often introduce ambiguity and should be avoided.
11. Words such as only, just, merely and others of a similar

nature should be used with care and moderation in writing statements.

12. Whenever possible, statements should be in the form of simple sentences rather than in the form of compound or complex sentences.
13. Avoid the use of words that may not be understood by those who are to be given the completed scale.
14. Avoid the use of double negatives.

APPENDIX C

Attitude Survey A

INSTRUCTIONS: Mark your honest feelings concerning Ornamental Horticulture as a field of study or profession. Do this by responding to the following statements. Use the answer sheet provided. Do not answer on the statement page or on your lecture answer sheets.

Some of the statements reflect agreeable attitudes or feelings; some reflect disagreeable attitudes or feelings.

You are to rate each statement according to HOW agreeable or HOW disagreeable an attitude or feeling it has on you. When you cannot decide one way or the other, you may mark undecided.

a	b	c	d	e
Strongly agree	Agree	Undecided	Disagree	Strongly disagree

1. It is great to work outdoors.
2. As a profession, Ornamental Horticulture adds much beauty.
3. The work in Ornamental Horticulture has a lasting benefit to people.
4. Trained monkeys can perform Ornamental Horticulture jobs.
5. Ornamental Horticulture makes people become depersonalized.

6. Working with a person's hands in Ornamental Horticulture gives enjoyment.
7. Fruit is a better gift for people.
8. Ornamental Horticulture is of little productive value to society.
9. Gardening can be done by homeowners without training.
10. Physical labor is too much in Ornamental Horticulture.
11. Ornamental Horticulture upsets nature.
12. Enjoyment in the beauty of life is provided by Ornamental Horticulture.
13. Ornamental Horticulture saves the environment.
14. Ornamental Horticulture provides for a beautiful world.
15. It is enjoyable work.
16. Ornamental Horticulture causes people to kill flowers.
17. Floral goods make people feel good.
18. It is unnecessary to society.
19. Ornamental Horticulture as a profession takes too much time.
20. Ornamental Horticulture is a luxury.
21. Ornamental Horticulture communicates gifts of nature.
22. Accomplishments in Ornamental Horticulture are very rewarding.

APPENDIX C

Attitude Survey B

INSTRUCTIONS: Mark your honest feelings concerning Ornamental Horticulture as a field of study or profession. Do this by responding to the following statements. Use the answer sheet provided. Do not answer on the statement pages or on your lecture answer sheets. Write your student number on the answer sheet.

Some of the statements reflect agreeable attitudes or feelings; some reflect disagreeable attitudes or feelings.

You are to rate each statement according to HOW agreeable or HOW disagreeable an attitude or feeling it has on you. When you cannot decide one way or the other, you may mark undecided. Note the responses under each letter.

a	b	c	d	e
Strongly agree	Agree	Undecided	Disagree	Strongly disagree

1. Ornamental Horticulture manipulates nature.
2. Ornamental Horticulture is unaware of new concept.
3. The conditions of work are bad.
4. Ornamental Horticulture makes people work long hours.
5. Ornamental Horticulture is for dummies.
6. Ornamental Horticulture is a trivial field of study.
7. There is satisfaction in Ornamental Horticulture from visual accomplishments.
8. Ornamental Horticulture is a very rewarding profession.

9. Ornamental Horticulture is very pleasing work.
10. People are able to work in open air.
11. Ornamental Horticulture creates a psychological lift.
12. Ornamental Horticulture is very stimulating to people.
13. Ornamental Horticulture is great for the health.
14. Ornamental Horticulture destroys essential land use.
15. Ornamental Horticulture creates beauty.
16. Working with a living plant form is satisfying.
17. Regardless of the field, plants are satisfying.
18. Living plants make living more bearable.
19. Ornamental Horticulture is for stupid people.
20. A degree is not worth anything for the job.
21. Ornamental Horticulture gives little prestige to people.
22. Unskilled workers can do the jobs in Ornamental Horticulture.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Aleamoni, Lawrence M. Why Is Grading Difficult? National Association of Colleges and Teachers of Agriculture Journal, March 1979. Vol. XXIII, No. 1.

2. Amthor, W.D. Is There Merit in a Pass-Fail Grading System? Paper presented at the meeting of the Americam Industrial Arts Association, Minneapolis, April 1968.

3. Barker, Donald G. Development of a Scale of Attitudes Toward School Guidance, Personnel and Guidance Journal, June 1966.

4. Berk, R.A. Determination of Optional Cutting Scores in Criterion-Referenced Measurement. Journal of Experimental Education, 1976. Vol. 45, 4-9.

5. Block, J.H. Mastery Learning, Theory and Practice. Holt, Rinehart and Winston, Inc., 1971.

6. _____. Schools, Society and Mastery Learning. Holt, Rinehart and Winston, Inc., 1974.

7. _____. Student Learning and the Setting of Mastery Performance Standards, Educational Horizons, 1972, pp. 183-191.

8. _____. The Effects of Various Levels of Performance on Selected Cognitive, Affective and Time Variables. Unpublished Ph.D. dissertation, University of Chicago, 1970.

9. _____. Criterion-Referenced Measurements: Potential. School Review, February 1971, pp. 289-298.

10. Campbell, D.T. and Stanley, J.C. Experimental and Quasi Experimental Designs for Research, 1973. Rand McNally College Publishing Company.

11. Carlson, J.G. and Minke, K.A. Fixed and Ascending Criteria for Unit Mastery Learning, Journal of Educational Psychology, Vol. 67, No. 1, 1975, pp. 96-101.

12. Carroll, John B. Problems of Measurement Related to the Concept of Learning for Mastery, Educational Horizons, 48, No. 3, 1970, pp. 71-80.

13. Cooley, W.W. and Lohnes, P.R. Multivariate Data Analyses, 1971. John Wiley and Sons, Inc.

14. Ebel, R.L. Criteria Referenced Measurements: Limitation. School Review, February 1971, pp. 282-288.

15. Educational Development at Michigan State University, Analysis of the Mastery Instructional Model, No. 5, Spring 1973, Office of the Educational Development Program, Michigan State University.

16. Edwards, Allen L. Techniques of Attitude Scale Construction. Prentice Hall, Inc., Englewood Cliffs, NJ, 1957.

17. Finn, J.D. and Mattsson, I. Multivariate Analysis in Educational Research. Applications of the Multivariance Program, 1978. International Educational Services, Chicago.

18. Foth, Henry D. A Mastery Learning Program in Soil Science, Journal of Agronomic Education, Vol. 2, November 1973, pp. 65-68.

19. _____. Improving Learning with Mastery Learning, Department of Crop and Soil Science, Michigan State University, East Lansing, MI 48824

20. Glaser, Robert. Adapting the Elementary School Curriculum to Individual Performance, Proceedings of the 1967 Invitational Conference on Testing Problems, Princeton: Educational Testing Service, 1968, pp. 3-36.

21. Gray, William M. A Comparison of Piagetian Theory and Criterion-referenced Measurement, Review of Educational Research, Spring 1978, Vol. 48, No. 2, pp. 223-249.

22. Hambleton, R.K., Hutten, L.R., and Swaminathan H. A Comparison of Several Methods for Assessing Student Mastery in Objective-Based Instructional Programs. Journal of Experimental Education, 1976, 45, 57-64.

23. Hambleton, R.K.; Swaminathan, H.; Algina, J.; Carlson, D.B. Criterion-Referenced Testing and Measurement: A Review of Technical Issues and Developments, Review of Educational Research, Winter 1978, Vol. 48, No. 1, pp. 1-47.

24. Hapkiewicz, Walter G. and Foth, W.G. Can All Students Learn What They Are Taught? Educational Development Program Report, Michigan State University, East Lansing, MI, No. 35, January 1973.

25. Harris, T.C., Kiefert, J.T. and Darby, M.D. Attitudes Expressed by Students Toward a Beginning Course in Educational Psychology, Journal of Educational Research, Vol. 62, No. 8, April 1969.

26. Johnson, J.J., Gnagey, W.J. and Chesbro, P.M. The Effectiveness of Applying the Concept of Mastery to the Teaching of Educational Psychology. Paper presented at the meeting of the American Educational Research Association, Minneapolis, March 1970.

27. Malott, R.W., Editor. Research and Development in Higher Education: A Technical Report of Some Behavior Research at Western Michigan University, Summer 1971.

28. Mehrens, William A. and Lehmann, Irvin J. Measurement and Evaluation in Education and Psychology, Holt, Rinehart, Rinehart, Winston, Inc., 1973.

29. Merrill, M.D., Barton, Keith and Wood, L.E. Specific Review in Learning a Hierarchical Imaginary Science, Journal of Educational Psychology, 61, 1970, pp. 102-109.

30. Millman, J. Passing Scores and Test Lengths for Domain-Referenced Measures. Review of Educational Research, 1973, Vol. 43, No. 2.

31. Neidt, Charles O. and Hedlund, Dalva E. The Relationship Between Changes in Attitudes Toward a Course and Final Achievement, Journal of Educational Research, Vol. 61, No. 2, October 1967.

32. Sherman, J.G. Application of Reinforcement Principles to a College Course. Paper presented at the annual meeting of the American Educational Research Association, New York, 1967.

33. Wentling, T.L. Mastery Versus Non-mastery Instruction with Varying Test Item Feedback Treatments, Journal of Educational Psychology, Vol. 65, No. 1, 1973, pp. 50-58.

34. Wentling, T.L. and Erickson, Richard C. Measuring Student Growth, Allyn and Bacon, Inc., 470 Atlantic Avenue, Boston, MA, 1976.

35. Wofford, J.C. and Willoughby, T.L. Attitudes and Scholastic Behavior, Journal of Educational Research, Vol. 61, No. 8, April 1968.

General References

1. Borg, W.R. and Gall, M.D. Educational Research, 2nd. Edt., 1971, David McKay Company, Inc.

2. Cox, D.R. Planning of Experiments, 1958. John Wiley and Sons, Inc.

3. Finn, J.D. Multivariance, A Fortran IV Program, Version 4, June, 1968. Department of Educational Psychology, State University of New York at Buffalo.

4. Little, T.M. and Hills, F.J. Statistical Methods in Agricultural Research, 1972. University of California, Agricultural Extension, AXT-377.

5. Steel, R.G.D. and Torrie, J.H. Principles and Procedures of Statistics, 1960. McGraw-Hill Book Company, Inc.

6. Wilson, F.R. and Sookpokakit, S. Multivariate Casebook: A Guide for Use of Finn's Multivariance Program in Processing Univariate and Multivariate Analysis of Variance, Covariance, and Repeated Measures Designs, Occasional Paper No. 31, August, 1978. Office of Research Consultation, College of Education, Michigan State University, East Lansing, Michigan.