1980

Evaluation of self-paced, mastery and traditional instruction in a university anatomy and physiology course using comparative and attribute by treatment interaction designs

Thomas A. Stinard
Iowa State University

STINARD, THOMAS A.

EVALUATION OF SELF-PACED, MASTERY AND TRADITIONAL INSTRUCTION IN A UNIVERSITY ANATOMY AND PHYSIOLOGY COURSE USING COMPARATIVE AND ATTRIBUTE BY TREATMENT INTERACTION DESIGNS

Iowa State University PH.D. 1980

Evaluation of self-paced, mastery and traditional instruction in a university anatomy and physiology course using comparative and attribute by treatment interaction designs

by

Thomas A. Stinard

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Professional Studies in Education
Major: Research and Evaluation

Approved:

Signature was redacted for privacy.

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1980

TABLE OF CONTENTS

CHAPTER I. INTRODUCTION

The last decade has seen increasing interest in self-paced, mastery instruction manifested in the use of such approaches as the Personalized System of Instruction (Keller, 1968) and Mastery Learning (Bloom, 1968). While the traditional lecture method is in no immediate danger of being supplanted, the self-paced, mastery alternative has won support from a growing number of educators. Two reasons for this popularity have been the inherent appeal of the instructional philosophy to the practicing teacher and the positive results from evaluative research.

Although there is no formal philosophy for self-paced, mastery instruction, the basic tenants hold that: (1) individual learning differences vary greatly from one student to another; (2) nearly all students can and should master all of the necessary material in a given course of study; and (3) if mastery is expected of diverse students, the instruction must recognize and provide for individual differences, for example, in rate of learning, instructional preferences, previous knowledge, learning styles, and student schedules. The philosophy clearly involves a more individualized

approach than the traditional lecture.

The evaluative research on self-paced, mastery instruction has been extensive. The findings are summarized in the next chapter. The present discussion is confined to the evaluation designs and to an apparent contradiction between the prevailing design and the philosophy of alternative instruction. Early evaluations were nonexperimental case studies where no control group was used. The typical report described the rationale and procedures for instruction and testing in some detail. Results consisted of descriptive data on student reactions and grades. Conclusions usually included opinions on how the new method compared with the status quo and a discussion of the advantages and implementation problems. The case study design was weak because the benefits of the new method above and beyond that of traditional instruction could not be measured. However, depending on the objectivity and purposes of the author-observer, this design was valuable at the developmental stages of alternative instruction.

The case study approach is still used, but the most common design is the comparative study. In this model a traditional and alternative teaching method are implemented in two reasonably equivalent groups and the

students are compared on cognitive and affective outcomes at the end of the term. The purpose of evaluation is evident in the following quote by Nation and Roop (1975, p. 108):

> . . .teachers are faced with countless possibilities (instructional techniques), most of which are attractive and have some support from well controlled research projects. How, then, is an individual interested in improving educational practice, going to make a decision? . . . of central concern. . . is which mastery technique best fulfills the basic needs of the students, i.e., which procedure results in the best classroom performance.

The comparative design seeks to determine the best instructional procedure by comparing the average performance of one group with the average performance of the other. This process of averaging across the group is, however, contrary to the instructional philosophy of self-paced, mastery teaching because it ignores individual differences in course performance. Instruction which provides for individual learning differences should be evaluated with a procedure which also recognizes individual differences (Latta, Dolphin, and Grabe, 1978).

Such evaluation methodology has been suggested by Cronbach and Snow (1977, p. 1), and in the following quote they qualify the search for the best method:

> The educator continually devises and applies new instructional treatments, hoping for

> improved results. He seeks the best method of instruction for a given purpose. Since learners differ, the search for generally superior methods should be supplemented by a search for ways to fit the instruction to each kind of learner. One can expect interactions between learner characteristics and instructional method. Where these exist, the instructional approach that is best on the average is not best for all persons.

This design, called attribute (or aptitude) by treatment interaction research, is particularly appropriate in the evaluation of alternative instruction because it empirically tests the possibility that an instructional procedure is more effective with certain individuals than with others. Cronbach first proposed the attribute by treatment interaction design in 1957, but significant applications in evaluating self-paced, mastery instruction were not made until the mid 70's. Both the methodology and research findings are reviewed in the next chapter.

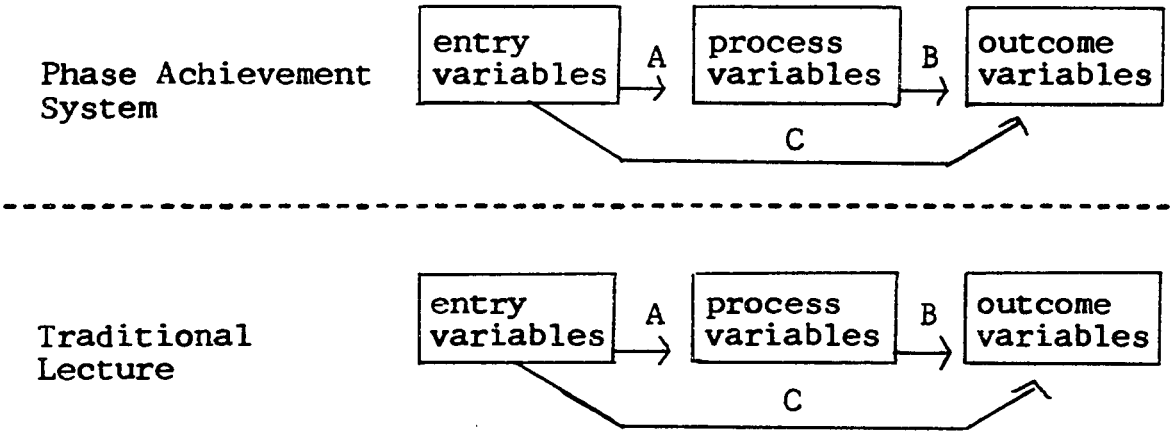
Another approach in evaluating alternative methods has focused on study behavior, or what is termed here, process variables. The purpose has been to more fully understand how students learn under instructional methods. At a simple level, process variables have been treated like outcomes; that is, compared between groups. For example, several researchers have found that total study time reported by students is higher in alternative than traditional instruction (see reviews by Kulik et al.,

1974, and Robin, 1976). Other investigators have examined the relationships between the distribution of study time over the term (particularly cases of procrastination and acceleration) and course outcomes and student characteristics. One set of process variables which has received little attention is the allocation of study time to various learning resources.

Purposes

One example of a self-paced, mastery alternative is the Phase Achievement System developed by W. D. Dolphin and colleagues (this approach is described in Chapters II and III). As part of a project funded by the National Science Foundation Comprehensive Assistance to Undergraduate Science Education Program, the Phase Achievement System (PAS) and a traditional lecture (TRAD) were implemented in two large sections of an introductory zoology course (total n about 300). The purposes of the present study are: (1) to evaluate PAS and TRAD using a comparative design; (2) to study the relationships between individual student differences, study patterns, and course outcomes; and (3) to evaluate PAS and TRAD using an attribute by treatment interaction design.

The specific objectives can be more easily explained with the help of an organizational model of variables. The following scheme was adapted from the ideas of Astin and Panos (1971) and Bloom (1976):



For each instructional method there are three categories of student variables (a glossary of all variables is located in Appendix C):

1. Entry variables are measures of individual student differences or attributes that students bring to the course. These can be subset into cognitive variables such as scores on college entrance exams, background in science, and grade point average, and into noncognitive variables such as locus of control, test anxiety, and preference for instructional methods.

2. Process variables in the present research are descriptors of study behavior; for example, the self-reported time spent studying and allocation of time to the textbook and study guide.

3. The outcome or dependent variables are measures of student attainment of course goals. The

> major outcome is course achievement measured by score on a comprehensive final exam given to both groups. There are also several attitudinal outcomes which address student opinion of the instructional procedures.

The arrows in the model represent the linear and nonlinear relationships between the variables. For example, Arrow C symbolizes the relationships between the several entry variables and the outcomes.

The first purpose is to evaluate PAS and TRAD with a comparative design. The main question is, Which instructional method is more effective in terms of student achievement and attitudes? Process (study) variables will also be compared between methods, but before the outcome and process variables can be validly compared, the entry variables must be compared to insure that PAS and TRAD students were reasonably equivalent as they entered the course of instruction. In terms of the model, the means of the variables in the boxes will be compared across the dotted line.

The second purpose is to predict and explain achievement, attitudes, and study behavior in the course by investigating the relationships (correlations) between the variables in the overall group. This phase of the research does not address the effectiveness of PAS and TRAD, but rather the influence of individual differences

upon outcomes. The major question is, Which types of students show the greatest achievement and which hold the most positive attitudes? In terms of the model, the focus is on the arrows. Specific objectives are listed below:

1. To understand the variables by examining the intercorrelations among entry variables, among process variables and among outcomes.

2. To determine the degree of influence of student charcteristics and study patterns on final exam score (Arrows A and B in the model).

3. To determine if the pre- and post-questionnaire data (personality measures, instructional preferences, and process variables) contribute to the prediction of final exam score above and beyond the variables obtained from university archival records. Because the administration and follow-up of questionnaires is time consuming, it is important to determine the predictive utility of variables derived from the questionnaire.

4. To determine the degree of influence of student characteristics and study patterns on attitudes toward the course by identifying the major correlates of the affective outcomes (Arrows B and C).

5. To determine the degree of influence of student attributes on study patterns by identifying the major correlates of the process variables (Arrow A).

The third purpose is to evaluate PAS and TRAD with the attribute by treatment interaction (ATI) design. Earlier it was argued that instruction, particularly self-paced, mastery instruction, should be evaluated with

a design which considers individual differences. The ATI design accomplishes this by joining the comparative approach (Purpose 1) with the individual differences or correlational approach (Purpose 2). The main question is, Which instructional method is more effective for certain types of individuals? In terms of the model, the relationships symbolized by Arrows A and B for PAS will be compared with those for TRAD. The statistical methods are explained in the next chapter. In addition to identifying ATI's, the process variables will be examined in an attempt to understand how the interactions occur.

The nature of Purposes 2 and 3 is exploratory or investigatory. Quite a large number of variables will be explored as predictors of outcomes and as attributes which might interact with instructional method. The testing of a set of a priori hypotheses was viewed as too confining at the present stage of inquiry into PAS. The investigation, however, is not done without expectations based on a review of relevant literature (see Chapter II). The main limitation of exploratory research is that when a large number of variables is considered there is a greater probability of finding significant relationships and interactions due to chance happenings. Consequently,

there is a greater need for replication in this type of investigation than in hypothesis testing research. Some internal replication will be attempted; that is, convergent evidence from several sources within the study will be sought. Strong interpretation and inference must await external replication with other students.

CHAPTER II. DEFINITIONS AND LITERATURE REVIEW

In this chapter findings from comparative and aptitude by treatment interaction studies of self-paced, mastery instruction are reviewed. The chapter begins with definitions for alternative and traditional instruction. The Personalized System of Instruction (PSI) is explained because it has been by far the most frequently evaluated of the self-paced, mastery approaches. The Phase Achievement System is also introduced and its similarities and differences with PSI are discussed. Comparative evaluation literature is summarized followed by a review of statistical methodology for ATI design and evaluative results.

Definitions

Alternative or self-paced, mastery instruction includes a number of approaches described in the literature. These methods are typically, but not necessarily, implemented in large enrollment introductory college courses of nine to fifteen weeks' duration. Course content is determined by the instructor or departmental policy. As defined here, alternative approaches share four characteristics.

1. Learning objectives are carefully selected and explicitly stated. The objectives for a course are organized into meaningful, sequential study units (anywhere from eight to thirty units). Proponents of alternative instruction believe that it is not only ethical, but pedagogically necessary, to tell students what is expected in precise terms.

2. Some form of mastery is required. Some procedures require a pre-set level of mastery within a unit (depth mastery); others require that a certain number of units be mastered (breadth mastery). Course grades are based on absolute levels of mastery making it possible for the entire class to receive high grades.

3. There is a conscious effort to relate test items to learning objectives. Testing is done by study unit and is therefore more frequent than in traditional courses. The purposes of testing are to document mastery and to identify areas of nonmastery.

4. More decision making is required of the student than in traditional lectures. Some type of student pacing is employed to provide for individual differences. Well prepared or fast learners can proceed through a course quickly, and students can adjust their study effort in accordance with other demands on their time. Students are also more actively involved in selecting learning materials and methods to use.

Traditional instruction (TRAD), as defined here, consists of large group lectures held several times a week with or without discussion sections. TRAD is not simply the absence of the four characteristics discussed above, but rather, more a matter of degree. For example, instructional objectives are used in TRAD but they

usually are less explicit. As with alternative instruction the goal of TRAD is mastery of course content, but an a priori level of mastery is not required. Students compete with each other for the highest grades, and the main purpose of testing is to normatively assign grades. Multiple choice exams are administered two to four times per term on an established schedule without provisions to retake tests to correct deficiencies. All students progress through the material at the same pace which is set by the instructor.

The most popular example of alternative instruction is the Personalized System of Instruction (Keller, 1968). In PSI perfect or near perfect mastery (90% to 100% correct) of a unit is required before a student advances to the next unit. Grades are based on the number of units mastered. Unit tests are available on demand, administered by proctors, and scored immediately. The type of test items are multiple choice or short essay and the responses are oral or written. If a unit is not mastered, the student (1) is told to restudy missed concepts; (2) is directed to specific learning resources; or (3) is tutored in deficient areas. Direct group instruction is minimal. The few lectures that are given are designed to motivate self-study.

PSI is self-paced in that once given the learning objectives and resources, the student decides when and how to distribute his time among the learning resources. There is the requirement that a unit must be mastered before the next one can be attempted. In some recent versions, deadlines for certain units have been imposed to reduce procrastination.

The Phase Achievement System (PAS), developed by Dolphin et al. (1973), is a teaching procedure which can be used as either an alternative or supplement to traditional instruction. The procedures are described in detail in Chapter III; the present discussion compares and contrasts PAS with PSI. Mastery for a unit of study is set in the 50 to 60% correct range which is considerably less than perfection. Mastery on all units is required, and grades are based on the average unit performance. In PSI, depth mastery is required and breadth mastery determines the grade. In PAS, breadth mastery is required and depth mastery determines the grade. Both methods are designed to promote full depth and breadth mastery of the subject material.

In PAS, tests over all units are available at test sessions scheduled approximately every two weeks. Students may take any combination of unit tests at a

test session. Test items are multiple choice, responses are machine scored, and results are available the following day. Test results are not personally discussed with students. The student decides if and when to restudy and retake a unit test to attain mastery or to improve a unit score. Lectures are regularly scheduled[1] and cover most of the content of the tests. Attendance is encouraged but optional. A study guide is available. PAS is self-paced in the sense that students can read assignments ahead of lecture schedules and take unit tests. The major constraint on self-pacing is that tests are offered every two weeks rather than several times a week as with PSI. The economics of academia impose this limitation, however, not the educational philosophy.

Review of Comparative Studies

A large number of comparative evaluations of PSI-type approaches have been conducted, and these have been reviewed by several authors (Ryan, 1974; Klauw and Plomp,

---

[1]With regular lectures PAS is considered a supplement to traditional instruction. It can be used without lectures. An "alternative" PAS course is scheduled for the near future; a comprehensive library of videotapes will be available to students on demand, providing greater opportunity for self-pacing.

1974; Robin, 1976; Kulik, Kulik, and Smith, 1976; Kulik and Jaksa, 1977). Two of the more comprehensive reviews are briefly summarized here.

Robin (1976) located 39 studies which compared TRAD with alternative instruction (27 were classified as PSI, the remainder were modifications of PSI). The outcome measures were scores on multiple choice or essay exams. The achievement of PSI students significantly exceeded that of TRAD students in 30 of the comparisons. One comparison significantly favored TRAD, and in the other eight there were no significant differences. The average score difference between the 39 PSI and TRAD groups was 9%. Robin also reviewed seven retention studies with intervals ranging from two months to two years. In all comparisons, the PSI group scored significantly higher than the TRAD group with the average difference on the follow-up measures being 13%.

Kulik and Jaksa (1977) came to similar conclusions after their review of 39 evaluations comparing PSI with TRAD. There was some overlap in studies between this and the Robin review. Thirty-four of the 39 comparisons on final exams revealed significant differences in favor of PSI; the remainder showed no differences. The average

difference in test scores across all studies was 13%. In nine studies which investigated retention over 3 to 60-week intervals, all differences were statistically significant and in favor of PSI. The average difference was 24%, indicating that the long term benefits of PSI were greater than the immediate effects.

The positive effects of PSI were evident in students' attitudes as well as in academic performance. Kulik, Kulik, and Smith (1976) found that in eight out of nine reports, PSI students rated their instruction significantly higher than did TRAD students. Robin (1976) reported that 14 of 16 attitudinal comparisons favored PSI-type instruction while the remainder showed no significant differences.

The comparative research has been decidedly in favor of PSI. This literature comparing PSI with TRAD is quite a contrast to the literature comparing the discussion method with lecture as reviewed by Dubin and Taveggia (1968)--of 88 studies, half favored the discussion method and half favored lecture, and most of the differences were not statistically significant. No doubt the Hawthorn effect operated in some of the PSI groups, but not nearly to the degree which would explain the consistently positive results.

The purpose of the comparative evaluation design is to determine whether PSI or TRAD is the better method as measured by student achievement. The answer, for the present, is PSI. This approach to evaluation, however, is not complete. The ATI design explores an additional question which is important to any instructional method and particularly relevant to self-paced, mastery instruction--Which students achieve higher with alternative instruction and which with TRAD? Even though PSI students on the average exceed TRAD students, it is still relevant to ask which types of students do particularly well under PSI and which types do significantly better under TRAD procedures.

The remainder of this chapter will focus on the ATI evaluation design discussing first ATI methodology followed by a review of ATI research.

Attribute by Treatment Interaction Methodology

The most acceptable statistical designs which provide evidence for attribute by treatment interactions are analysis of variance (ANOVA) and the regression approach (Cronbach and Snow, 1977). Both test three sources of influence upon the outcome variable: the effect of the

level of the attribute variable; the effect of the treatment or instructional method variable; and the effect of the interaction between the attribute and treatment. If the attribute is a nominal variable, the ANOVA and regression designs are identical. If the attribute is of a continuous type, which is usually the case, it is artificially divided into two or more levels when ANOVA is used. For the regression model, the attribute retains its continuous nature. The interpretation of a significant interaction for both approaches is that students at certain levels of an attribute do significantly better on an outcome under one treatment compared to another. For the regression model, a significant interaction is also another way of saying that the slope of the best fit regression line for one treatment group is significantly different from that of another treatment group.

In addition to the ANOVA and regression designs, there are several approaches which either purposefully or incidentally provide information about ATI's but the methodology is flawed. These are called quasi-ATI designs here; the evidence they provide is weaker, but they are included in the literature review because there are so few well designed studies of alternative instruction.

One quasi-ATI design is the correlational approach

in which a continuous attribute and outcome are correlated in a TRAD and alternative group. The larger the difference between the correlations the more evidence there is that one instructional method results in greater achievement for certain students than the other method. For example, if ability were highly correlated with course performance in TRAD but unrelated in PSI, then it is the brighter students who excel in TRAD. But in PSI some brighter and some less bright excel. Given this situation there is evidence that PSI is differentially beneficial for bright students and TRAD is differentially beneficial for the less bright.

It is possible to statistically test for differences in correlations (see Hays, 1963), but none of the studies located actually did so. The correlational approach is a good preliminary step in describing differential relationship, but for statistical tests the difference in slopes in the two treatments should be performed, not differences in correlations (Cronbach and Snow, 1977). This is because the slope is measured in the actual units of the attribute and outcomes and provides a more natural, accurate description of the relationship. Correlations, on the other hand, are computed on standardized variables where means and variances are artificially equated.

Differences in slopes are frequently paralleled by differences in correlations, but not always, as in cases where attribute and outcome variances differ between instructional treatments.

Another quasi-ATI approach is the use of multiple t-tests. An attribute, say scholastic aptitude, is measured for all students and PSI and TRAD are implemented in two comparable sections. The attribute is categorized (e.g. high, medium, and low), and four t-tests are done which compare outcomes (all PSI versus all TRAD, high PSI versus high TRAD, medium PSI versus medium TRAD, and low PSI versus low TRAD). Interpretations are then offered. If, for example, all the tests were significant, the conclusion would be that PSI was more effective than TRAD at all levels of aptitude. If none of the tests were significant except, say, the high PSI group exceeded the high TRAD group, then there is some (weak) support that PSI is differentially beneficial for the bright students. The ANOVA or regression approach described earlier is superior to this type of analysis because multiple t-tests inflate the level of significance (see Hays, p. 275 and p. 471, 1963), less advantage is taken of the potential degrees of freedom yielding less power than ANOVA, and there is no definitive tests of the interaction between

attribute and method.

The above four designs (ANOVA, regression, correlation, and multiple t-tests) require two comparable groups; one taught with TRAD and one with self-paced, mastery instruction. Indirect evidence can also be gained about ATI's which use only one group. Correlations between an attribute and outcome for TRAD sections can be compared to similar correlations found in alternative groups. Such comparisons are usually weak because of dissimilar students, course content, and measures, but if there are consistent trends across several studies, then valuable evidence is realized. There is a type of one-group design which should not be used. Cross and Semb (1975) and Semb (1976) reported negative correlations between pre-test scores (attribute) and gain scores (post-test minus pre-test scores used as an outcome) in several PSI courses. Noting that the lower pre-test students gained more than the higher pre-test students, they interpreted the correlations to mean that PSI was differentially beneficial for low pre-test students. Actually, the negative correlations are explainable solely on the basis of regression to the mean and part to whole score confoundment (see Linn and Slinde, 1977).

The research which is pertinent to ATI-type questions varies greatly in quality of design. In the following section, ATI, quasi-ATI, and one-group studies have all been included with an effort to interpret the findings with only as much certainty as the design allows.

Review of Attribute by Treatment Interaction Studies

As with the comparative evaluations, nearly all of the ATI literature has investigated PSI-type procedures as the alternative instruction. Instances are noted where the alternative and traditional methods deviated from the description of PSI and TRAD given earlier in this chapter. The review is organized by attribute.

Many studies investigated scholastic aptitude as an interacting attribute with the hypothesis that PSI is differentially beneficial for low aptitude students. This expectation is based on the tenant of mastery learning that nearly all students can master the course material. Assuming that is true, one would expect aptitude and performance to be unrelated in alternative instruction (see Bloom, 1976, and Kulik and Kulik, 1976). Initial study of the ATI research on scholastic aptitude, however, was very contradictory. Some researchers reported that

low aptitude students benefitted most from self-paced, mastery methods, others reported no differential benefit, and a few even reported greater benefit for high ability students (see brief reviews by Cross and Semb, 1976, Kulik and Kulik, 1976, and Morris and Kimbrell, 1977). Further study of the ATI literature, including quasi-ATI and one-group studies, has revealed that if a distinction is made between general aptitude (for example, cumulative grade point average, high school rank, and college entrance exam scores) and specific content related achievement (for example, pre-tests and performance in similar subject areas), then many of the contradictions are rectified. The research was reviewed with aptitude and prior achievement as separate attributes.

Another student characteristic which has received attention in the ATI literature is internal-external locus of control. To a lesser degree, test anxiety, preference for instruction, and achievement motivation have been studied, and these attributes, in turn, are reviewed.

General aptitude as an interacting attribute

One-group studies There are many studies reporting strong relationships between aptitude and performance in traditionally taught college courses. Rather than review

all of them, results from life science courses at Iowa State University are presented since they are most applicable. Reports from the Iowa State University Testing and Evaluation Center (1975, 1976, 1977, and 1978) reported correlations for freshmen who took Biology 101 during the fall quarters of 1974 to 1977. The sample sizes varied from 900 to 1100. For ACT composite scores, the correlations with course grade ranged in a tight band from .55 to .62. For high school rank (HSR) and the Minnesota Scholastic Aptitude Test (MSAT), the ranges were -.48 to -.54 and .42 to .51 respectively.

Wagner (1977) obtained nearly identical results for the same course offered in Winter Quarter, 1977. The subjects were 230 freshmen and upperclassmen, and the correlations for grade with ACT, HSR, and MSAT were .59, -.52, and .51. The correlation for cumulative grade point average was very high, .80. In Biology 103, a second quarter course, Wagner found correlations for ACT, HSR, MSAT, and GPA to be .50, -.49, .36, and .69 (n = 130, Winter Quarter, 1977). Bennett (1970), in an early study on achievement in Biology 101, correlated HSR and MSAT with grades and exam scores (n = 857). Results for grades were commensurate with those just cited; for exam scores the correlations were significant but

lower (HSR, r = -.22; MSAT, r = .35). Only Wagner (1977) reported correlations within sexes. There were no notable sex differences except MSAT correlated somewhat higher for males than females.

The second half of the one-group studies concerns the relationship between general aptitude and performance in alternatively taught courses. Because of the dearth of studies it was necessary to extend the scope beyond life sciences at Iowa State University. In the studies which were located, the instructional method was PSI and the content area was psychology. Wood and Wylie (1975) found a dramatic relationship between college GPA and performance on unit quizzes (n = 147); students with GPA's from 1.00 to 2.00 earned on an average only 13% of the total course points. Those with GPA's from 2.01 to 3.00 and 3.01 to 4.00 earned 47% and 81%. Arguing that PSI procedures should be particularly beneficial for lower ability students, they found the results discouraging. Further investigation yielded similar findings. ACT composite scores correlated .59 with final exam scores for one group (n = 34). The correlations for SAT scores reported by Nazzaro et al. (1972) were lower but significant. With a sample size of 93, SAT verbal and SAT math correlated .29 and .27 with

final exam scores. GPA was considerably higher, .50.

There were five other studies which reported upon the relationship between GPA and course performance. Allen et al. (1974) obtained significant differences between students in the top and bottom halves of GPA (total n = 88) on course grades, performance during oral quizzes, and scores on a multiple choice final exam. There were no differences on an essay exam. Franklin (1976) did not report correlations but did find GPA to be the most predictive of course achievement among a number of cognitive and affective variables. The data presented by Henneberry (1976) showed that GPA was highly related to course grade for both slow and fast starting students (n = 304). The author argued that because lower GPA students started at a slower pace and slow starters performed poorer, "self-pacing may be very inappropriate for the poorer student." While the conclusion sounds reasonable, an important question remains--How do lower GPA students perform in TRAD? The research just reviewed indicates that they, too, achieved at lower levels in comparison with high GPA students.

Davis (1975) found a significant relationship between GPA and test performance in two versions of PSI; one required 50%, the other 100% mastery before a student

could progress to the next unit of study. The only one-group study which did not find a significant relationship between aptitude and achievement in an alternative course was reported by Calhoun (1975). The correlation between GPA and final exam scores for 231 PSI psychology students was not significant.

There is good agreement among the one-group studies that general aptitude is highly related to achievement in traditional courses. The evidence also points to similarly high relationships in alternative courses. Of the aptitude measures, SAT was the lowest, but nonetheless, it correlated significantly in the studies located. As stated earlier, the one-group studies are valuable in answering ATI questions only if they agree with well designed two-group studies.

Two-group studies As part of a larger study, Latta, Dolphin, and Grabe (1978) reported slopes and correlations between MSAT and grades in two comparable sections of introductory biology. In the section taught with the Phase Achievement System (n = 191), the correlations were .44 for females and .56 for males; for the TRAD group (n = 194), they were .59 and .50. The differences in correlations and slopes were not substantial for either sex.

Morris and Kimbrell (1972) also used a correlational approach and found that SAT composite scores correlated .61 with final exam scores for a PSI section (n = 39) and .75 for a TRAD section (n = 37) of psychology. The authors concluded that PSI was not particularly beneficial for low ability students in comparison with high ability students. Several years later (1977), they eliminated data for several subjects for a reanalysis (ANOVA) and concluded that PSI was differentially beneficial for low SAT students. Their reasoning on excluding students, however, was faulty. They argued that PSI students who did not take the final exam (n = 2) and who did not complete all of the PSI units (n = 7) should be eliminated because they were actually "non-PSI" students, and their inclusion biased the analysis. They were correct on the first point but not on the second. Students who were taught under PSI conditions must be considered PSI students in spite of their failures to do well. In reality, it is their exclusion, not inclusion, which biased the results of the reanalysis because when the poorer students were eliminated, the variance of the final scores in the PSI section was reduced, which in turn, lowered the correlation in the PSI group by an unknown amount.

Pairs of correlations were reported by several other

authors. Born and Davis (1974) found a correlation of .70 between unspecified college entrance exam scores and performance on a final exam in a PSI psychology course (n = 63). The correlation was nearly the same in the TRAD section (r = .72, n = 31). Ludwig (1977) reported no significant differences between correlations in a PSI and TRAD section. The correlations were between the school and college aptitude test and midterm scores. The difference in slopes was not tested. The correlations betweem SAT and final exams reported by VanVerth and Dinan (1974) and Kulik, Kulik, and Milholland (1974) were similar in PSI and TRAD sections.

Kulik and Kulik (1976) obtained a correlation between SAT and final exam scores of .63 (n = 151) for a PSI group but only .30 (n = 65) for the TRAD students. The sizes of the correlations indicated that PSI was more beneficial for the brighter than the less bright students, and this was verified through ANOVA. The authors reported a significant interaction between SAT and instructional method giving support to a finding which is contrary to the expectations of PSI proponents. The work of Schimpfhauser et al. (1974) has been interpreted by others (see Morris and Kimbrell, 1977, Cross and Semb, 1976, and Gindler et al., 1977) as evidence that PSI differentially

benefits higher ability students. The Schimpfhauser results are descriptive only, however, and are difficult to interpret because of the use of difference scores (actual minus predicted achievement scores). The authors themselves made no conclusions on differential benefit.

In three investigations, GPA was studied as an interacting attribute using the ANOVA approach. Jacko (1974) reported results on PSI and TRAD sections of a textile and clothing course (total n = 88). The measure of course performance was score on an unannounced examination. Both treatment and GPA main effects were significant but the interaction was not. Callis (1977) found essentially the same with her study. The treatments were television lectures/supervised laboratory (here considered TRAD) and independent study with a total sample size of 92 clothing design students. The course outcomes were measured by knowledge and application sections of a final exam. The interaction between GPA and method was not significant for either outcome.

DuNann and Weber (1976) briefly reported on a two-year follow-up study on students from a contingency managed course (similar to PSI) and a TRAD course in introductory psychology. Eighty-six (35%) of the original students were located and completed an achievement test

over the course material. Preliminary analyses indicated no evidence of differential dropout. Current college GPA was trichotomized and was included as a main effect together with instructional method in an ANOVA analysis design of variance. Using the follow-up test as a dependent measure, they found significant or near significant main effects but no GPA by treatment interaction. That is, long term retention of material was equally superior in the PSI offering across the three levels of GPA. They also used analysis of variance with original final exam score as the outcome. The GPA by treatment interaction was significant but this analysis was not appropriate. It should have been done on all original students using GPA at the time the course was taken.

In summary, almost all of the evidence from the two-group studies indicated that general aptitude and instructional method do not interact, that alternative instruction is not differentially beneficial to low or high ability college students. The one-group studies strongly supported this finding. It is concluded that students who are high in general aptitude achieve more regardless of whether they are taught with traditional or alternative approaches as defined earlier. The conclusion does not, of course, preclude the possibility

that, with modification, alternative procedures will negate the powerful influence of general aptitude. In fact, several authors cited here have set such goals in redesigning instructional procedures.

Specific achievement as an interacting attribute

One-group studies In reviewing the literature, specific achievement was distinguished from general aptitude. Specific achievement is defined as a measure of previous knowledge directly related to the subject matter being taught in contrast to general aptitude which reflects previous academic performance across a variety of subject areas. Examples of specific achievement are performance on a test or in prerequisite courses.

Bloom (1976) presented extensive evidence in his review that previous achievement is strongly related to future achievement. This was the case across several traditional subject areas and age levels. His studies, however, were confined to the elementary and secondary school levels. The most relevant research cited by Bloom came from a study by Flanagan (1964) where the correlations between prior and subsequent achievement in TRAD high school biology courses were in the .60's. Several large studies at the college level were located which examined the effects of taking high school science

upon achievement in TRAD life science courses.

Johnsten (1967), for example, found significant differences in biology gain scores among TRAD students who had taken (1) no high school chemistry, (2) at least one high school science course, and (3) chemistry in both high school and college. He also reported a significant relationship between high school physics and subsequent achievement in college biology. Bennett (1970) reported a significant relationship between taking high school chemistry and scores on a comprehensive biology exam but no relationship between high school biology and the exam. On the basis of data gathered over two years, Tamir (1969) concluded that most students enrolled in biology were severely penalized by not taking biology and/or chemistry in high school.

In a correlational study at Iowa State University, Wagner (1977) found significant relationships between the number of high school science semesters (biology, chemistry, and physics) and grades in two TRAD biology courses. For Biology 101 ($n = 307$) the correlation was .28. The correlations for males and females were comparable. For Biology 103 ($n = 169$) the correlation for the males was somewhat higher than for the females (.48 versus .24). The correlations for high school science in the Wagner study

were all significant, but they were lower than for general aptitude. This may have been because the measure reflected only the number of courses taken, not the level of performance. Significant but modest relationships between prior and subsequent achievement in traditional college biology courses were also reported by Szabo and Feldhusen (1971) and Hopper (1968).

Few studies of alternative instruction which reported relationships between specific achievement and course performance were located. Only one was found where life science was the content; Szabo and Feldhusen (1971) implemented an audio-tutorial botany course (n = 215) and found neither high school science GPA nor CEEB Science scores significantly related to course grade. Hess (1974) presented raw data on 45 PSI psychology students. Calculations on the data revealed a nonsignificant correlation of .14 between pre-test and post-test scores. Calhoun (1975) reported near zero correlations for major (psychology or nonpsychology) and number of psychology courses taken. The outcome was score on a final exam for a PSI psychology course (n = 231). In the above three studies, it did not appear that restriction of range of the outcome variables would explain the lower correlations.

Two-group studies Born et al. (1972) presented

descriptive data showing that PSI "narrowed the gap" between students who were high and low on initial achievement. The sample sizes were very small, however, and their design was weak as pointed out by Fernald and DuNann (1975). In a replication, Fernald and DuNann randomly assigned psychology students to one of two groups. One section received TRAD teaching throughout the semester (n = 91), the other received TRAD for the first half of the semester and PSI for the last half (n = 91). On the basis of test scores covering the first half of the semester, the students were classified as high or low achievers. An analysis of variance on gain scores[1] (final minus midterm scores) was computed. Both main effects were significant, but the interaction was not. In other words, PSI students significantly exceeded TRAD students in gain scores, the low achieving TRAD students gained about the same as the high achieving TRAD students, and the low achieving PSI students gained about as much as high PSI students.

Latta et al. (1978) examined a different type of measure for specific prior achievement. They reported

---

[1]The use of gain scores did not present a serious problem here because regression and confoundment effects operated equally in the TRAD and PAS sections.

upon the relationship between the number of high school science semesters (biology, chemistry, physics) and course grades in a TRAD section and a section taught with the Phase Achievement System (PAS). There was little evidence of a differential relationship; for the PAS and TRAD males the correlations were .28 (n = 99) and .17 (n = 115), for PAS and TRAD females the correlations were .23 (n = 101) and .30 (n = 92).

Pascarella (1977, 1978) reported results from three investigations designed specifically to explore attribute by treatment interactions. In these studies the students chose to take introductory calculus in either a PSI or TRAD section. The self-selection ruled out a strict experimental design, but Pascarella provided evidence that the two groups were equivalent on important entry characteristics in all three studies. The measure of specific entry achievement was score on a math placement exam. In the first investigation (total n = 94), the math pre-test was trichotomized and analysis of variance was done with the dependent variable being performance on a problem solving final exam. Both main effects were significant as well as the interaction. The low pre-test PSI studients did significantly better than the low pre-test TRAD students. In the second study with 60 PSI

and 188 TRAD students, the math pre-test was kept as a continuous variable and a regression approach was used. Again the two main effects (pre-test and treatment) and the interaction term contributed significantly to the explanation of final exam (similar to the exam described in the first study). The third study replicated all earlier results with 53 PSI and 57 TRAD calculus students. In the last studies the math pre-test correlated .43 and .55 with final exam for the TRAD section. For the PSI group the correlations were significantly lower, .22 and .23.

Cross and Semb (1976) also provided evidence that specific achievement interacts with instructional method. They found that low PSI achievers, as determined by earlier achievement tests, were significantly higher on test scores than low achievers in a TRAD section; this was not the case for high achievers. Cross and Semb used the multiple t-test approach which weakened, but did not negate, their conclusions. Two other researchers have reported data which support the position that PSI-type courses favor students who are weaker in relevant prior knowledge. Austin and Gilbert (1973) drew this conclusion on the basis of descriptive data for PSI and TRAD physics students.

Gindler et al. (1977) also presented descriptive statistics indicating that PSI algebra students who were low on a pre-test outperformed comparable TRAD students. At the high end of the pre-test, the TRAD and PSI were about equal. Enough information was provided to compute approximate correlations between the pre-test and final exam. They did show a fairly large difference between treatments (PSI, $r \simeq .3$, $n = 262$; TRAD, $r \simeq .6$, $n = 146$).

In summary, Fernald and DuNann (1975), who used a true ATI design (ANOVA) with adequate sample sizes, did not find a significant interaction between prior achievement and instructional method. Pascarella (1977, 1978), on the other hand, did in three separate studies. He also used sizable samples and an appropriate design (regression). The treatments, of course, were not identical, but according to the descriptions, they contained all of the salient components of PSI. The negative results of Fernald and DuNann might have been because PSI was operational for only a half semester.

When consideration is given to the quasi-ATI designs, the bulk of the evidence indicated that specific achievement interacted with instructional method. The literature as a whole revealed that students who were less well prepared in specific content achieved higher under PSI

than TRAD conditions. Apparently PSI is remedial; it provides a setting which allows or encourages students to make up deficiencies in prerequisite knowledge. In traditional instruction, however, those who are deficient in prior knowledge tend to do poorly in subsequent achievement.

Internal-External Locus of Control as an interacting attribute

One-group studies Internal-External Locus of Control (IE) as defined by Rotter (1966) is the degree to which a person sees himself in control of his life and the events that influence it. Internals see themselves as exerting significant influence while externals tend to believe that events are determined by forces outside their control, such as fate, luck, and powerful others. IE was originally investigated as a variable which might refine the predictions made from social learning theory. It has also been explored as a possible factor in achievement with the expectation that internals would perform better than externals in an academic setting. The research has not born this out; numerous studies have found the Rotter IE scale and college GPA to be unrelated (Eisenman and Platt, 1968; Hjelle, 1970; Warehime, 1972; Gozali et al., 1973; Prociuk and Breen, 1973).

When achievement in science courses was used as the outcome rather than GPA, the results were mixed. Wagner (1977) reported weak relationships between IE and final grades in two TRAD biology courses at Iowa State University. For 164 males and 154 females in Biology 101, the correlations were -.16 and -.09. With IE scored in the external direction, the correlations were in the expected direction; that is, externally oriented students earned lower grades. The correlation for the males was significantly greater than zero but, nontheless, quite low. The correlations for Biology 103 were not significant ($r = .05$, $n = 99$ males; $r = -.14$, $n = 61$ females).

Lipton (1976) found IE to be related to neither midterm nor final exam scores for 114 TRAD physics students. In a TRAD psychology course, Boor (1973) reported near zero correlations for females ($n = 61$), but for the males there was a significant relationship ($r = .31$, $n = 55$). These results contrast with correlations reported by Massari and Rosenblum (1972) for TRAD psychology students. They found a positive correlation (opposite of that expected) for females ($r = .27$, $n = 43$) and essentially no relationship for males ($n = 50$).

Considering both the GPA and course achievement studies, it appears that the Rotter IE scale has little

value in predicting cognitive outcomes in traditionally taught college courses.[1] Several authors have suggested that the relationship has been weak and inconsistent because of the nature of traditional instructional (see Allen et al., 1974; Rotter, 1975; and Parent et al., 1975). With normative grading and fixed lecture and testing schedules, TRAD does not appear to be the type of instruction which would allow the IE construct to clearly manifest its relationship with academic achievement. On the other hand, researchers have argued that in self-paced, mastery settings, there is more potential for student control of outcomes (e.g. the option to retake tests) and IE should be more highly related to course performance.

Three one-group studies were located which addressed the relationship between IE and performance in PSI courses. Allen et al. (1974) found a high correlation between IE and grade in an upper level psychology course ($r = -.46$, $n = 88$). In fact, the IE to grade correlation was higher than the GPA to grade correlation. Keller et al. (1978) did not find a relationship ($r = .04$, $n = 138$ introductory psychology students). Their criterion was total number of course points and it was apparent that there was no serious restriction of its range.

---

[1]Stronger relationships have been found for children and adolescents. See Bar-Tal and Bar-Zohar (1977).

Johnson and Croft (1975) reported a near zero correlation between IE and grade for 179 upper level psychology students. Little can be concluded, however, because the design called for about half of the students to receive some external progress monitoring by proctors causing the instructional method to be more traditional than PSI. Another problem was the restricted variance of course grade; with 86% of the students receiving a grade of "A", the correlation may well have been artificially reduced.

Two-group studies Several two-group IE studies were located, but unfortunately, none of the alternative methods were strictly PSI procedures; they did share some components of PSI.

Three of the investigations presented evidence of an interaction in which the alternative method was differentially beneficial for internally oriented students. The best controlled study was done by Daniels and Stevens (1976). Students were randomly assigned to either a TRAD method or a contract plan. TRAD students in introductory psychology were required to attend lectures, complete assigned text readings, and take weekly quizzes. Grades were assigned on a normative basis. Under the alternative method the students contracted for grades by specifying proficiency levels for quizzes over seven units

and projects in subject areas of personal interest. Lectures were not required. Testing was self-paced, and unmastered units and inadequate projects could be made up with additional work. Students in the top and bottom quartiles of IE were included in an analysis of variance of final exam scores (n = 86). Neither the IE nor treatment main effects were significant, but the interaction was. The internal contract students did better on the final than the internal TRAD students, while the external contract students did poorer than the external TRAD students.

This finding was supported by two quasi-ATI studies. Eilersen (1972) reported that IE was not related to test scores in a TRAD psychology course but was significantly and negatively related in an individualized learning system (total n = 116). Hohn et al. (1977) reported similar results when comparing TRAD with self-paced instruction in educational psychology. The sample size was quite low--only 36 students in all. Neither Eilerson nor Hohn formally tested for an IE by treatment interaction.

Two researchers did not find the expected interaction. Somon (1976) implemented a self-paced and a teacher-paced educational psychology course (total n = 204). IE was

dichotomized at the median and an ANOVA revealed no significant IE by method interaction. Reynolds and Gentile (1976) hypothesized a significant interaction such that internals would perform better under mastery than TRAD assessment conditions while the reverse would be true for externals. The subjects experienced both methods. The hypothesis was tested with undergraduate and graduate educational psychology students using the ANOVA approach. For the 76 undergraduates the interaction was not significant. The interaction was significant for the 44 graduate students but in the wrong direction. It is important to note that the instructional method was considerably different from PSI. The student was required to take post-tests over four units at fixed times; if he did not pass at the 80% level, "he was apprised of his area of weakness . . . and helped to learn the material." The unit was retaken until passed. It does not appear that there was more opportunity for student control under the mastery assessment condition than the TRAD; in fact, the treatment may have encouraged a fatalistic attitude because it forced eventual mastery.

Conclusions from the IE research are difficult to draw. Not only were the results mixed but the alternative treatments varied somewhat from PSI. Of the two studies

employing true ATI designs (Somon, 1976, and Daniels and Stevens, 1976), one did not find a significant interaction and one did. The one-group studies were also mixed. Most of the studies found IE and achievement to be unrelated in TRAD classes, but in PSI classes one found a high correlation and another did not. The literature as a whole leans toward supporting the interaction hypothesis, but too few well-controlled studies have been conducted to make any conclusion.

Other variables as interacting attributes

The attributes reviewed in this section have not received much attention in the ATI research on alternative instruction, but on the basis of theory and logical reasoning, they hold potential as interacting variables. The attributes include test anxiety, student preference for instruction, and achievement motivation.

One of the more popular measures of test anxiety has been the Test Anxiety Questionnaire (TAQ, sometimes called the Test Anxiety Scale) of which there are several versions. The following review is confined to the 21-item version described by Sarason (1958). Correlations between TAQ and course achievement in traditionally taught courses have been reported in areas of biology and psychology. The correlations with final grade were mildly

to moderately negative in the introductory biology courses at Iowa State University (Biology 101: n = 164 males, r = -.22; n = 155 females, r = -.37. Biology 103: n = 90 males, r = -.14; n = 94 females, r = -.20). The correlations were comparable in introductory psychology when TAQ was related to final exam scores (Alpert and Haber, 1960: n = 40 males, r = -.35. Carrier and Jewell, 1966: n = 94 males, r = -.32; n = 57 females, r = -.43).

Given the nature of testing in self-paced, mastery instruction (students are allowed to retake tests, there is immediate feedback of results, and the students don't compete with each other for high grades), the negative relationship between TAQ and achievement might well be moderated in alternative courses. Appealing as the argument may be, little relevant research was located. The only one-group study located supports this hypothesis. Allen et al. (1974) found that TAQ was not related to course grade, score on an essay test, or score on a short answer exam (n = 98 males and females in a PSI psychology course). The only two-group study (PAS versus TRAD) found no evidence of an interaction. The Latta, Dolphin, and Grabe (1978) research indicated sex differences between methods (PAS: n = 99 males, r = -.01; n = 101 females, r = -.35. TRAD: n = 115 males, r = -.09; n = 92 females,

r = -.38).

Student preference for instruction as a determinant of performance has been studied by several researchers. Logically speaking, one would expect that students who prefer alternative over traditional instructions would achieve higher under alternative conditions. Cronbach and Snow (1977) reviewed the research on this hypothesis and concluded that "the evidence discourages the romantic view that self-selection of the instructional diet pays off." About one half of the studies they examined were laboratory studies, and none employed true ATI designs or self-paced, mastery instruction *per se*.

A search for more relevant literature on preference did not fare much better. Looking outside the domain of ATI research on self-paced, mastery methods, two studies concurred with the Cronbach and Snow conclusion (Parent *et al.*, 1975, and Dorsel, 1976), and one did not (Gaynor and Millham, 1975). In the latter study, the authors measured preference for various teaching and testing conditions and randomly assigned about 120 psychology students to one of six instructional methods (discussion versus laboratory versus lecture by frequency of testing). They found that the degree of concordance between preference for instruction and actual

instruction received correlated very highly with performance on course exams. According to their multiple regression table, concordance explained 46% of the variance in performance beyond that explained by GPA.

Because alternative methods allow retakes of tests, it would seem reasonable to expect that motivation would be highly related to achievement and perhaps more highly related under alternative than TRAD conditions. There is some evidence to support an interaction. Pascarella (1977) reported a well controlled, regression ATI study. The measure of achievement motivation was taken from the Stern Activities Index and the subjects were 47 PSI and 47 TRAD students in calculus. Final exam scores and satisfaction with the course were used as dependent variables, and significant interactions were found for both. PSI differentially benefited the more motivated students. The correlations between motivation and performance were .51 and .02 in PSI and TRAD.

The present investigation uses an abbreviated version of the Resultant Achievement Motivation Scales consisting of a Fear of Failure Scale and a Hope for Success Scale (see Chapter III). Two one-group studies were located on these scales and both were conducted at Iowa State

University in TRAD biology courses. The correlations ranged from zero to .26 with no consistent patterns across scales or genders. No research is available on the relationship of these scales to performance in self-paced, mastery instruction.

The literature does not afford conclusions on the attributes reviewed in this section: for test anxiety, the results are few and mixed; the bulk of the literature on preference for instruction indicates that preference does not interact, but most of the treatments were not strictly self-paced, mastery; and the research on Resultant Achievement Motivation is too sparse. Clearly, more research is needed on these attributes.

Summary of Literature Review and Relevance to Present Study

The great majority of the foregoing literature employed PSI versions of self-paced, mastery instruction. As noted in the definitions section of this chapter, PSI and PAS are theoretical and operational exemplars of self-paced, mastery instruction. Differences were also noted, but it is believed that the PSI literature is a viable history upon which to guide the present research and speculate upon its outcomes.

Comparative literature

The preponderance of evidence from the comparative evaluations supports the conclusion that self-paced, mastery instruction is statistically and educationally superior to traditional instruction. Although the intent of PAS is to promote mastery of course material and to provide for individual learning differences by modifying, not replacing, traditional testing and grading procedures, it is expected that students taught with PAS will, on the average, learn more and report greater satisfaction than students taught with traditional instruction.

Attribute by treatment interaction research

Several general observations about the ATI literature are warranted before summarizing the substantive findings. Although the ATI approach was proposed more than 20 years ago (Cronbach, 1957), it is still very much in its infancy. Well-executed research is the exception, not the rule, particularly for self-paced, mastery instruction. A distinction was drawn between true ATI, quasi-ATI, and one-group designs, but an attempt was made to synthesize the results from all three types and to weight each proportionately to its credibility.

Invariably the dependent variable has been cognitive performance. Only one study was located which investigated

an ATI in the prediction of student attitudes toward instruction (Pascarella, 1977). Granted that course achievement is the primary goal of higher education, there is no reason for the exclusion of affective outcomes or study behavior. Another neglected area in the self-paced, mastery literature is empirical research to explain how or why ATI's take place.

The attribute which has received the most attention has been general academic aptitude. A number of researchers have included not only the global measures of past performance but also measures of more content specific achievement in the category of general aptitude. It was found that the literature was more consistent if a distinction was made between the two.

There was a high level of agreement among the studies that general aptitude did not interact with instructional treatment. Aptitude was highly related in both traditional and self-paced, mastery instruction. It is expected that the measures of aptitude in the present study--college grade point average, high school rank, and scores on the ACT and MSAT entrance exams--will not interact with instructional method.

There was less agreement on specific achievement as an interacting variable than with general aptitude,

but the bulk of evidence indicates that this attribute did interact with instructional method. Students with poorer backgrounds in specific, content related areas tended to learn more in alternative than TRAD instruction, and students with better backgrounds learned more in TRAD settings. Only one measure of content specific background is included in the present study, the number of semesters of science (biology, chemistry, and physics) taken in high school. It is expected that this variable will interact with instructional method.

The literature on IE locus of control is mixed. When consideration is given to all of the one-group, true ATI and quasi-ATI studies of PSI and related self-paced, mastery instruction, the evidence leans toward the presence of an interaction with PAS favoring the internal students and TRAD favoring the external students. In the present investigation, IE and two subscales based on factor analysis (see Chapter III) will be explored as interacting variables, but no expectations are offered because of the mixed findings in the literature.

The ATI research on self-paced, mastery instruction for test anxiety, preference for instruction, and achievement motivation is too meager to generate solid expectancies. The inclusion of these attributes is based

more on logical speculation than on empirical evidence. The Test Anxiety Questionnaire, Resultant Achievement Motivation Scale, and preference for alternative testing procedures represent the variables. Their measurement is explained fully in Chapter III.

CHAPTER III. METHODOLOGY

A complete description of the subjects, course of instruction, instructional and testing procedures, measurement of variables, and data collection are included in this chapter. The statistical methods are explained in the results chapters.

Subjects and Course of Instruction

The subjects were students enrolled in two sections of Zoology 155 during Spring Quarter, 1977, at Iowa State University. Assignment of subjects to instructional methods was not strictly random. The students registered for one of the two sections about three to six weeks prior to Spring Quarter. They were not aware that one section would be taught with the PAS method. Both sections met for a total of 150 minutes per week, but one met two times a week and the other met three times. The PAS method was randomly assigned to the section with three lectures per week. The comparability of the groups is explored in Chapter IV.

Most of the analyses were done on data from the 300 students who took the final exam common to both sections.

This included 36 PAS males, 108 PAS females, 47 TRAD males, and 109 TRAD females. There was a disproportionately large number of females, particularly in view of the all undergraduate ratio which is about one female to two males. About one-half of the subjects were freshmen and one-third were sophomores. Descriptive statistics on other characteristics are available in Chapter IV. Only about one-third of the students were science majors (see variable called MAJOR in Table 4 of Chapter IV). There is evidence that the students as a group were above the national average in aptitude (see ACT and HSR in Tables 5 and 6, Chapter IV). The average students took four and one-half semesters of high school science (HSSCI in Table 5, Chapter IV).

The subjects can be further characterized by describing the course and type of students who take it. Zoology 155 is an introduction to physiology and anatomy of human organ systems. The course was designed for students in the College of Home Economics--about one-third of the Spring 1977 enrollment were home economic students--but it is open to all students. In recent years, many pre-nursing and physical education students have taken the course. Students are advised to take a college biology course or a rigorous biology course in high school before

enrolling in Zoology 155. Lecture sections range in size from 100 to 250 students. There are no accompanying discussion sessions though there is a separate laboratory course which several students take. The course is offered three times a year with about 1000 students taking it per year.

Instructional Procedures

The learning conditions for the sections were as similar as possible except that PAS was used in one section. Dr. Joyce Emery taught both sections. Spring Quarter was her third quarter teaching Zoology 155. The instructional core for both sections was the study guide. This detailed outline of the course content was written by a departmental committee headed by Dr. Yola Forbes. The study guide was organized into nine units or phases corresponding to the organ systems (body organization, skeletal, nervous, muscular, circulatory, digestive, respiratory, urinary and reproductive systems). Each phase was further divided into subphases and finally into study questions. Both lectures and tests drew heavily from the guide and study questions. The students in both sections were encouraged to purchase and use the guide which was available at the local bookstore for the cost

of printing. The textbook was Principals of Anatomy and Physiology (Tortora and Anagnostakos, 1975).

The Phase Achievement System was implemented as an alternative testing and grading procedure in the present study. Regularly scheduled lectures were given as in TRAD. In order to receive a passing grade the student was required to master each of the nine phases. Mastery was demonstrated by passing phase tests which were offered at six test sessions held outside of the lecture period every two weeks. If a student did not pass a phase or wanted to improve a phase grade, he or she could prepare for and retake a parallel test at another test session. No penalty was assessed for retaking tests; that is, the highest score was used in determining mastery and the grade for a phase. There were few limitations on retakes. Only tests over Phases 1, 2 and 3 could be taken at the first test session, and only Phases 7, 8, and 9 could be taken at the last session. With a total of six test sessions, Phases 1, 2, 3, 7, 8, and 9 could be attempted up to five times, and Phases 4, 5, and 6 up to six times.

The tests were randomly generated by phase from a total item bank of about 2000. A conscious effort was made to include only those items which were clearly

associated with the study guide. There were 18 multiple choice items on the tests for each phase, and these were inspected by the instructor for errors and adequacy of coverage. Because of computer support, the results from the most recent test session and cumulative records were available the day after the session. Mastery on a phase was defined as 10 or more correct out of 18 items (more than 56%). Letter grades for each of the nine phases were assigned according to the following schedule: less than 10 items correct = F; 10 or 11 items correct = D; 12 or 13 = C; 14 or 15 = B; 16, 17 or 18 = A. There was, of course, some arbitrariness in setting the criteria. Previous experience with PAS in biology courses indicated that the schedule was reasonable.

Eighty percent of the course grade was based on the phase grades, about 9% for each phase. The remaining 20% was based on performance on a comprehensive final exam composed of 80 multiple choice items. These items were not used on the phase tests. The final could not be retaken, and it was included in PAS only to provide a common measure of course achievement between the two sections. The use of a final exam is not necessary in PAS, and they were not used in previous PAS sections. An incomplete grade was given only if seven or eight

phases had been passed and if the final had been taken. The incomplete grade had to be made up within the next year by passing the unmastered phases. If less than seven phases were mastered by the end of the quarter, the student received a failing grade.

The main differences between PAS and TRAD in Spring 1977 were the testing procedures and grading policies. In TRAD there were no testing options. The students were required to take the two 60-item midterm exams and the 80-item common final at the pre-set times. Retakes were not allowed. The midterms, which shared many items with the PAS phase tests but none with the final, contributed 40% each in determining the course grade. The final was weighted 20%, the same as in PAS, to insure that motivation for taking the final was about the same for each section. Total course points were calculated for each TRAD student after weighting the test scores, and grades were normatively assigned; that is, a pre-set percentage of students received A's, B's etc. after examining the curve of the distribution for natural breaks.

Measurement of Variables and Data Collection

Entry variables

Student entry variables came from two sources,

university archival records stored on computer tapes and a pre-questionnaire which was specifically constructed for this study.

The archival tapes contained a wide variety of individual student data. Selection of the variables was guided in part by the literature and also by previous experience in college science education. These data were hand checked against student files from the Student Admissions Office on a sample of about 25 students. The only inconsistencies were in ACT scores as explained below. Abbreviations for the archival variables are listed below with full descriptions and notes on rates of usable data. The response rates were calculated on the basis of the 300 students who completed the final course examination in both sections because the majority of analyses employed this group of students.

1. TOTHRS--Total hours was the number of college quarter hours earned as of the beginning of the experimental quarter, Spring Quarter, 1977. Credit transferred from other colleges and universities and credit earned by examination were included. TOTHRS was available for all students.

2. MAJOR--This dichotomous variable divided students into science and nonscience majors. The following majors were classified as science: Animal Science, Biochemistry, Biophysics, Dairy Science (Pre-Vet), Entomology, Fisheries and Wildlife Biology, Plant Pathology, Food and Nutrition, Bacteriology, Biology, Botany, Chemistry, Computer Science, Earth Science, Mathematics, Metallurgy, Physics, Psychology, Sociology, Statistics, Zoology,

Preparation for Human and Veterinary Medicine, and all majors in the College of Engineering. There were other science majors offered in the various colleges, but they are not listed because none of the students in the samples had declared them. Undeclared Science and Humanities students were assigned to the nonscience group. The classifications were made on the basis of expert judgment and there was some arbitrariness. The department course requirements were examined in marginal science majors. Data for the science-nonscience variable were complete for all cases.

3. HSR--High school rank was a percentile score which reflected academic standing relative to a student's graduating class. A HSR of 20 means the student was exceeded by 20% of his/her class in cumulative high school grade point average. The higher the score, the lower the standing. Of the 300 students who took the final exam, 272, or 91%, had usable HSR scores. Missing values were due mainly to students who transferred from other universities.

4. GPA--Grade point average was based only on credit hours earned at Iowa State University. There were eight students who had no credit hours; GPA's for these students were treated as missing values, yielding a response rate of 97%.

HSR and GPA were cognitive achievement variables based on relatively long term performance in actual academic settings. The next two variables (MSAT and ACT) reflect cognitive achievement as assessed by timed college entrance tests.

5. MSAT--The Minnesota Scholastic Achievement Test is a shortened form of the Ohio State University Psychological Test. There are three types of items (vocabulary, analogies, and reading comprehension), but only one score is reported. It is administered routinely to all entering ISU students except transfer students. MSAT scores were available for 80% of the students.

6. ACT--The American College Testing Program is a standardized entrance exam consisting of four subtests (English Usage, Math Usage, Social Studies Reading, and Natural Sciences Reading). The ACT is administered on a voluntary basis, primarily to college-bound high school juniors and to some seniors. The scores of the ACT are typically less current than the MSAT scores. The ACT is more content or achievement oriented than the MSAT which is a more homogeneous measure of verbal aptitude. Only the composite scores (based on the four subtests) were available from the archival computer tapes. The original response rate for ACT was only 67% on the computer tapes, but a hand search of printed files located an additional 32 students which raised the response rate to 78%.

7. HSSCI--High school science is the sum of the number of semesters of biology, chemistry, and physics taken in high school. It was decided to combine all science courses instead of using just biology because most students took two semesters of biology and the variance was low. Blank values on the computer file presented a problem because it was not known if they represented zero semesters or missing values. It was decided to treat blanks as missing if there were blanks for all subject areas and the student had transferred from another university. Otherwise, blanks were treated as zero. A visit with an admissions officer and a hand check on several students supported this decision. General science and health courses were not counted in these science subject areas. HSSCI was 91% complete.

The pre-questionnaire (Pre-Q) was composed of locally constructed items and published personality scales. The scales were selected on the basis of research with the Phase Achievement System at ISU in 1974 (Latta, Dolphin, and Grabe, 1978) and the literature reviewed in Chapter II. The Pre-Q was administered during the third lecture period

of the quarter for both sections. The purpose of the questionnaire was briefly explained and the students were asked to read and sign a written consent form (see Appendix A). Telephone follow-up of nonrespondents continued for three weeks. Optical scan answer sheets were used. They were visually checked for poorly made responses and corrected if needed. For the PAS section, 139 of the 144 students who took the final exam completed the PRE-Q for a response rate of 96%; 140 of the 156 TRAD students responded (89%).

The 14 locally constructed items on the Pre-Q (Items 1 through 13 and Item 66, see Appendix A) addressed preference for instructional methods, interest in the subject matter, instructional philosophy, motives, and expectancies. The students were asked to agree/disagree on a nine-point Likert-type continuum except for Item 66. This item asked for expected course grade and was recoded (A = 9, B = 7, C = 5, D = 3, F = 1) so its response format would approximate the other items. It was decided to combine items into scales for the sake of parsimony and because variables based on several items are preferable to single item variables (see Nunnally, 1967, pages 56-58). There are several procedures for combining items to form a scale. One way is to sum items on the basis of item

content. Another way is to combine items on the basis of inter-item correlations by visually examining the correlations or factor analyzing. A decision was made to use factor analysis as a first step and then temper the scale construction by examining the content of the items.

The principal factor analysis program (estimated communalities along the diagonal of the correlation matrix) provided by the Statistical Package for Social Sciences (Nie et al., 1975) was used for the 14 locally constructed items. All students who responded to all 14 items (including several who dropped the course) were included in the analysis (n = 284). The number of factors was determined by plotting the eigenvalue by sequential factor number and examining the plot for breaks in the curve. The breaks were not dramatic; visual inspection indicated either two or three main factors. Both the two and three factor solutions were rotated with the Varimax procedures. Inspection of both matrices indicated that the two factor solution was best (Table 1). The following variables were formed using the factor results. A caveat is appropriate, however. Factor analysis does not necessarily identify important items; it simply identifies items which do and do not cluster together.

8. CONFID--Confidence. Items 4, 5, 6 and 66 loaded highly and cleanly on the first factor for both the two and three factor solutions. They

Table 1. Varimax rotated factor matrix (Pre-Q items[a])

	Item	Two Factor Solution F1	F2
4.	Since I have a good background in science, I expect to do well in this course.	.80	-.05
5.	I see no benefit in taking this course in biology, but I must since it is a requirement for graduation.	-.55	-.06
6.	I have always been interested in biology.	.64	-.01
66.	Expected grade in course (A = 9, B = 7, C = 5, D = 3, F = 1).	.59	.10
1.	I prefer independent study and discussion sections instead of lecture classes.	.10	.66
10.	I prefer to take courses in large lecture sections.	.05	-.56
9.	I do not plan to spend a lot of time studying for this course.	-.04	-.12
2.	I am responsible for determining my progress and grade.	.20	.15
3.	Students should be allowed to take tests when they are prepared.	-.02	.32
7.	My friends told me that this course has a reputation for being difficult	-.23	-.04
8.	A study guide helps a student organize his study effort.	.15	.21
11.	I prefer tests once a week rather than twice a week.	.01	.34
12.	Grades should be based on the curve rather than pre-set standards.	-.18	-.03
13.	If students can retake tests, academic standards are lowered.	.17	-.24

[a]Items abbreviated. For exact wording see Appendix A.

are listed first in Table 1. The items ask about expectancy of success, self-confidence, and interest in biology. The items sample from both the cognitive and affective domains. Further definition of CONFID is available by examining the variables with which it correlates. The correlations are discussed in Chapter V. CONFID was constructed so that higher scores were associated with more confidence and interest in the subject matter. Item 5 was subtracted from the sum of Items 4, 6 and 66. CONFID was treated as missing if more than one item was omitted. If a student responded to three of the four items, the average was calculated and multiplied by four to estimate the sum for all four items. If all four items were completed the simple sum was used. CONFID was 93% complete.

9. PLECT--Preference for lecture. Items 1 and 10 (see Table 1) emerged as strong correlates of the second factor for both solutions. Item 1 was subtracted from Item 10 to create the variable. Students who scored high on PLECT showed strong preference for lecture methods as opposed to small group recitation or independent study. PLECT values were used only if both items were completed. The response rate was 93%.

Only one item (Item 9) loaded highly on the third factor of the three factor solution. The content of the second and third highest loading items (Items 8 and 12) and Item 9 did not appear to measure a unitary construct. Therefore, only two variables were constructed from the factor analysis. The CONFID and PLECT factors were replicated in an analysis of 715 students who took an identical questionnaire the preceding quarter (Winter Quarter, 1977). These students were enrolled in introductory biology courses (ISU Biology 101 and 103) and

the course of this study (ISU Zoology 155).

10. PALTT--Preference for alternative testing. It was expected that a factor addressing preference for testing procedures would emerge, but as illustrated in Table 1, this was not the case. Items 3, 11, 12 and 13 asked about some aspect of testing. Rather than combine these items on the basis of content, the correlations among the items were first examined. The matrix of correlations based on 300 students is presented below.

Item	11	12	13
3	.17	-.02	-.24
11		.05	-.07
12			.11

The 3, 11 and 3, 13 couplets showed the highest correlations. The 3, 11, 13 triad was not a cohesive combination because the correlation between Items 11 and 13 was near zero. Consequently, Items 3 and 13 were used for PALTT; the items addressed self-paced testing and opportunity to retake course tests. Item 13 was subtracted from Item 3 yielding a higher score for those students who reported greater preference for alternative testing procedures. If either item was omitted, PALTT was treated as missing data. The response rate was 93%.

The remaining entry variables originated from the published personality scales included in the Pre-Q.

11. RAM--Resultant Achievement Motivation. The items of this scale were selected from a 26-item measure of achieving tendency proposed by Mehrabian (1969). Drawing upon the theories of Mehrabian, Latta (1975) attempted to identify two subscales: one scale to measure Hope for Success (tendency to be motivated by needs to approach success) and another scale to measure Fear of Failure (tendency to be

motivated by needs to avoid failure). Factor analyses of the male and female versions of the original Mehrabian scales identified four items for each subscale. The items employed a nine point agree/disagree response format.

a. HS--Hope for Success was constructed by summing Items 35 to 38 (see Appendix A) for males. For females, Item 36 was reflected (9 = 1, 8 = 2, 7 = 3 etc.) and summed with Items 35, 37 and 38. HS was calculated only if all four items were completed; the response rate was 92%.

b. FF--Fear of Failure was the sum of Items 39 to 42 for both sexes. It was assigned missing if any of the items were omitted. The response rate was 93%.

c. RAM--Resultant Achievement Motivation was calculated by subtracting FF from HS. Latta (1975) found the construct validity of the difference score to be higher than for the sum of the HS and FF scores. The term "resultant" refers to this subtraction operation. The higher the score on RAM, the greater the tendency of a student to be motivated by hope for success versus fear of failure. RAM was 92% complete.

12. TAQ--The Test Anxiety Questionnaire used in this study consisted of 21 items (Items 14 to 34, Pre-Q, Appendix A). The items were the same as those reported by Sarason (1958) except the true/false response format was changed to a nine point agree/disagree format in order to increase the variance of the scale. Items 28, 33 and 24 were subtracted from the sum of the other items. TAQ was assigned missing if a student omitted three or more of the 21 items. Otherwise, values of the omitted items were estimated by averaging. The response rate was 93%.

13. IE--The Internal-External Locus of Control Scale (Rotter, 1966) measured a generalized expectancy for external versus internal control of rein-

forcement. Internally oriented individuals perceive themselves as directly responsible for their successes and failures; externals tend to attribute their successes and failures to sources beyond their control, such as to luck, fate, or powerful others. Each of the 23 items (Items 43 to 65 of the Pre-Q, Appendix A) presented an external and internal statement and the respondent chose the one he/she thought was most true. The first statement of Items 43, 47, 48, 49, 55, 59, 61, 62 and 65 was the external option while the second statement of the other items was external. One point was given for each external statement chosen. Those scoring higher on IE were more external. IE was considered missing if there were three or more omitted items. The response rate was 91%. An earlier unpublished study showed that IE correlated nearly zero with course achievement in introductory biology at Iowa Sate Univeristy (see literature review in Chapter II). Several researchers have suggested that the relationship between IE and academic achievement is low because the IE addresses achievement in many contexts not just academic settings (see Johnson and Croft, 1975, and Lipton, 1976). For example, Kaemmerer and Schwebel (1976), Collins (1974), and Mirels (1970) factor analyzed the full IE scale and found factors reflecting internal-external control over political events. In an attempt to eliminate nonacademic items from the full scale, the 23 items were factored using the same methods described earlier for the 14 locally constructed items on the Pre-Q. The subjects were 666 students enrolled in ISU Biology 101 and 103 and ISU Zoology 155 during Winter Quarter, 1977. The eigenvalue by factor number plot indicated five main factors, two of which were considered most relevant to the present research.

14. IELUK--IE Luck. Seven items had relatively high factor loadings on the first principal factor after Varimax rotation. Two of the items were discarded because they loaded on two factors (Items 53 and 62 loaded on the third factor).

The remaining five items and loadings are listed below.

Item 47. (.34) 1. Without the right breaks one cannot be an effective leader. 2. Capable people who fail to become leaders have not taken advantage of their opportunities.

Item 51. (-.57) 1. Becoming a success is a matter of hard work, luck has little or nothing to do with it. 2. Getting a good job depends mainly on being in the right place at the right time.

Item 54. (-.53) 1. In my case getting what I want has little or nothing to do with luck. 2. Many times we might just as well decide what to do by flipping a coin.

Item 55. (.60) 1. Who gets to be the boss often depends on who was lucky enough to be in the right place first. 2. Getting people to do the right thing depends upon ability, luck has little or nothing to do with it.

Item 57. (.41) 1. Most people don't realize the extent to which their lives are controlled by accidental happenings. 2. There really is no such thing as "luck".

Items 51 and 54 were subtracted from the other items. Students who were high on IELUK believed that luck versus ability was the predominant source of their success and failure. If a student omitted more than one item, IELUK was treated as missing. The response rate was 92%.

15. IEACAD--The third factor was called IE Academic. There were six items which correlated highly with the factor; two were discarded because of multiple loadings (Items 53 and 62).

Item 46. (.34) 1. The idea that teachers are unfair to students is nonsense. 2. Most students don't realize the extent to which their grades are influenced by accidental happenings.

Item 50. (.40) 1. In the case of the well-prepared student there is rarely if ever such a thing as an unfair test. 2. Many times exam questions tend to be so unrelated to course work that studying is really useless.

Item 61. (-.48) 1. Sometimes I cannot understand how teachers arrive at the grades they give. 2. There is a direct connection between how hard I study and the grades I get.

Item 64. (.40) 1. What happens to me is my own doing. 2. Sometimes I feel that I do not have enough control over the direction my life is taking.

All of these items but the last related directly to locus of control in an academic context. IEACAD was constructed by subtracting Item 61 from Items 46, 50 and 64. High scorers on IEACAD tended to perceive their academic achievement as a result of the fairness of teachers and exams and accidental happenings as opposed to degree of effort. At least three of the four items had to be completed before IEACAD was calculated. IEACAD was 92% complete.

The items which loaded highly on factor one (IELUK) and factor three (IEACAD) were remarkably consistent in their content. This was the case for the other factors as well. They are described here briefly for the benefit of other researchers. Items 52, 65, 60 and 65 loaded cleanly on factor two and referred to control in politics and world affairs; factor five items (45, 48, 58 and 63) addressed the origin of friendship; Items 43 and 49 loaded highly on factor four and addressed perceived causes of misfortune. Contentwise, the factor four items are related to IELUK, but they were not included on that

subscale because the factor analysis indicated a lack of strong relationship.

Process and outcome variables

The items of the post-questionnaire (Post-Q) were used to form the process (study behavior) and outcome variables except score on the final exam. There were three parts in the 50 item questionnaire. The first part consisted of seven items and used a variety of response formats (see Appendix A). The items of the second part (opinions about the instructor) used a nine-point agree/disagree format. The Post-Q was administered during the last week of classes for both sections. A list of TRAD and PAS students who had taken the Pre-Q but not the Post-Q was then prepared, and at the final exam periods the lists were checked when students turned in their exams. When asked to complete the questionnaire most did so at the exam period. The response rates for the Post-Q (again based on those who took the final exam) were 85% for the TRAD and 93% for the PAS section. The rates of return when considering both the Pre-Q and Post-Q were 81% and 90% for the PAS and TRAD students. The response rates were felt to be very good considering the fact that the questionnaires were somewhat lengthy and participation was voluntary.

Three process variables were calculated from the Post-Q items. Each was based on only one item because combining several items would confuse the meaning of the variables.

1. HRSPERQ--Hours per quarter was the number of self reported study hours spent outside of class time per quarter. Item 2 (Appendix A) asked for the number of hours per week. It was recoded to the mid point value (for example, if the third response was selected, the value of 5.5 hours was used; for the sixth response, 12 hours was used) and multiplied by 10 (there were 10 weeks of instruction for both sections). HRSPERQ was 89% complete.

2. TEXT%--Item 5 was used for the percentage of the textbook that a student reported reading. The response values of 1 through 5 were retained, but the following interpolation provides an approximate percentage scale: 1 = 10%, 2 = 30%, 3 = 50%, 4 = 70%, 5 = 90%. The response rate was 89%.

3. SGUID%--The percentage of the study guide questions answered by a student was estimated with Item 6 and interpreted as for TEXT%. The response rate was 90%.

An attempt was also made to calculate self-reported hours of lecture attendance as a fourth process variable. Unfortunately, Item 3 (Post-Q, Appendix A) presented psychologically different scales for the TRAD and PAS students because PAS students met for briefer but more frequent lectures per week than the TRAD students. The difference in scale meaning became apparent when prorated lecture hours per quarter was calculated; the means were not significantly different, but the variances

and shapes of the PAS and TRAD distributions were so disparate that none of the transformations provided comparable distributions.

There were three affective or attitudinal outcome variables constructed from the Post-Q. Two of the variables were formed from factor analysis of the second part of the questionnaire (Items 8 to 39, Post-Q, Appendix A). Before factoring it was decided to remove three items asking about course resources and to analyze them as single items. These items were 14, 25 and 28 which addressed quality of the textbook, lectures and study guide. The remaining 29 items were factor analyzed as described for the 14 locally constructed items on the Pre-Q.

There were 266 PAS and TRAD students (Spring Quarter, 1977) who responded to all 29 items. The eigenvalue plot indicated two or four main factors; both solutions were rotated and examined. The items which loaded highly on the first two factors for both the two and four factor solutions were identical. Only a few items loaded on the last two factors of the four factor solution; therefore, only two attitudinal variables were constructed. Table 2 presents the rotated factor matrix for the two factor solution.

1. F1--Factor 1. The first nine items listed in Table 2 were judged to be reasonable candidates for

Table 2. Varimax rotated factor matrix (Post-Q items[a])

	Item	F1	F2
11.	Interest in biol. devel. to where I wanted to spend more time.	.70	.11
32.	Level of interest increased as result of course.	.79	-.00
35.	Course allowed me to pursue areas of personal interest.	.44	-.14
36.	Course stimulated my desire to take more biology courses.	.69	-.13
18.	This is one of the better science courses.	.67	-.04
39.	I would recommend that others take this course.	.74	-.02
26.	This course forced me to regard myself unable to comprehend biol.	-.45	.39
37.	I feel I have mastered course content.	.61	-.16
29.	Test results were useful in planning my study schedule.	.46	.01
9.	Tests were threatening.	-.37	.64
10.	There was too much emphasis on tests and grades in this course	-.34	.41
30.	Number of exams was not adequate to test my understanding.	-.6	.41
16.	The grade standards were too high.	-.18	.65
31.	This has been a difficult course.	-.26	.75
38.	My final grade will be limited because I lack science background.	-.19	.47
24.	I spent too much time on this course.	-.07	.49
22.	I had freedom to arrange my study schedule.	.26	-.44
8.	I felt I had to do all the assigned reading to do well.	-.04	.19
12.	This course contains a lot of busy work not related to content.	-.30	.12
13.	Cramming was the most effective means of obtaining a high grade.	-.16	.05
15.	Tests were an adequate measure of my knowledge.	.44	-.42
17.	I felt I had to answer all study guide questions to do well.	.07	.24
19.	If supplemental videotapes were available, I would use them.	.07	.31
20.	I felt I could determine my grade more in this course.	.35	-.32
21.	I adjusted my study habits according to test scores.	.31	.05
23.	Frequent attendance in class is essential to good learning.	.15	.28
27.	I am satisfied with overall organization and instruction	.39	-.24
33.	I would prefer to take tests at my own pace.	.05	.08
34.	Grades should be based on a "curve", not on pre-set standards.	-.10	.33

[a]Items abbreviated. For exact wording see Appendix A.

forming a variable. The judgment was made on the basis of magnitude of factor loadings, degree of multiple loadings, and the item content. For example, Item 15 was not chosen as an F1 item because it loaded on both factors. Item 26 was selected even though there was some multiple loading because its content complemented Item 37 which correlated highly on F1. The F1 items reflected a positive general attitude toward the course, perceived increase in interest in life sciences, and sense of mastery of the material. Contentwise, there were three somewhat distinct areas, but the results of the factor analysis revealed that all nine items clustered together. It would have been psychometrically unsound to split the F1 items into subscales. The F1 variable was constructed by subtracting Item 26 from the sum of the others. Higher scores on F1 indicated more positive attitudes. At least seven of the nine items had to be completed before F1 was calculated, yielding a response rate of 89%.

2. F2--Factor 2. The second group of eight items in Table 2 was judged as a good representative of the second factor. The items addressed testing and grading procedures and perceived difficulty in completing the course. Not all of the items about testing emerged on F2 (e.g. Items 15, 21, 29 and 33). One common denominator of most of the items appears to be the students' perception of fairness about the course. F2 was constructed by subtracting Item 22 from the sum of the other items. High scores on F2 were associated with negative attitudes toward the testing and grading practices and greater perceived course difficulty. If a student omitted more than two items, F2 was treated as missing. The response rate was 89%.

3. INSTR--Instructor. The variable measuring attitudes about the instructor was not formed on the basis of factor analysis. All eleven of the items in the third section of the Post-Q (Items 40 to 50, Appendix A) were simply combined. Items 40, 44, 46, 47 and 49 were subtracted from the sum of the others. Those scoring higher on INSTR rated the instructor higher. The variable was calculated if there were four or fewer omitted items. The response rate was 88%.

The remaining items in Table 2 either had low or multiple loadings. The reader should be aware that an item's loading had nothing to do with the importance of the item, only with whether it covaried with other items. Several of the items which did not load on F1 or F2 were relevant to understanding the PAS and TRAD instruction. They are discussed in the results chapters.

CHAPTER IV. COMPARATIVE EVALUATION RESULTS

The purpose of Chapter IV is to compare course outcomes of students taught under traditional conditions (TRAD) with those under PAS conditions. The major question is, Which instructional method is better in terms of learning the course material and in terms of attitudes toward the course. Comparisons are also made of study (process) variables. Preliminary analyses are made on differential dropout and initial comparability of the groups to validate comparisons made later on the outcomes. The subjects and methodology are described as the results are presented.

Differential Dropout

Of the 174 PAS and 178 TRAD students initially enrolled in Spring Quarter, 30 PAS (17.2%) and 22 TRAD (12.3%) students dropped the course. These percentages did not differ significantly ($df = 1, \chi^2 = 1.30$, $p > .05$), but the type of student who withdrew may have differed. Study of differential dropout is important because it can lead to nonequivalent groups which in turn can distort the results from comparative study of

outcomes. Also, such evidence can be informative about the nature of PAS and TRAD instruction and their effects on students.

There were a number of entry characteristics[1] upon which to compare the PAS and TRAD dropouts. The PAS and TRAD students who withdrew did not differ significantly in GENDER or MAJOR according to the tests in Table 3. The TRAD students dropped significantly earlier in the quarter. Seventeen PAS students officially dropped in the last four weeks in comparison with only five TRAD students. These figures suggest that a "wait and see" attitude was more prevalent in the PAS section compared to TRAD.

Data from archival records and the pre-questionnaire were not kept for students who dropped in the first two weeks because this period was viewed as a time of schedule rearrangement. This caused small sample sizes, but t-tests were calculated under the premise that weak information about dropout is better than no information. The results, presented in Table 4, indicate that those who dropped from TRAD differed little from PAS dropouts. Only one of the twelve characteristics tested was significant. That variable was IE Locus of Control, with

---

[1]Chi-square tests were selected for dichotomous variables and t-tests for continuous variables.

Table 3. Chi-square tests for relationship of SECT with GENDER, MAJOR, and time of withdrawal

	PAS		TRAD		χ^{2a}
	obs.	exp.	obs.	exp.	
GENDER					
male	13	13.3	10	9.7	0.02
female	17	16.7	12	12.3	
MAJOR					
nonscience	27	26.0	18	19.0	0.02
science	3	4.0	4	3.0	
Time of withdrawal					
Week 1 - 2	6	12.1	15	8.9	10.27*
Week 3 -10	24	17.9	7	13.1	

[a] χ^2 calculated with Yate's correction, df = 1.

*$p < .01$.

Table 4. t-tests for mean SECT differences on student entry variables (students who withdrew after the second week)[a]

Variable	n	$\bar{x}$	s[b]	t[c]
ACT	10/4	24.0/22.0	4.83/7.07	0.62
MSAT	14/4	38.5/35.3	13.2/14.9	0.67
GPA	23/6	2.39/2.42	0.66/0.91	0.10
HSR	16/5	23.1/19.6	19.0/17.9	0.37
TOTHR	23/6	87.7/81.8	42.2/50.2	0.29
HSSCI	15/5	4.00/4.60	2.59/2.41	0.46
TAQ	13/4	72.8/72.3	23.8/15.6	0.04
RAM	13/4	5.00/-.25	9.29/11.7	0.93
PALTT	12/4	4.58/1.75	2.68/4.79	1.51
CONFID	13/4	9.85/13.8	5.84/6.13	1.16
PLECT	13/4	-3.2/-1.8	4.71/4.65	0.52
IE	13/4	11.3/15.8	3.08/2.63	2.58*

[a]For n (sample size), $\bar{x}$ (sample mean), and s (standard deviation) the value for the PAS students is to the left of the slash; that is, PAS/TRAD.

[b]Hartley's F_{MAX} Test indicated common SECT variances for all variables. Therefore, the standard error of mean difference was estimated by pooling the sample variances.

[c]Two-tailed tests.

*$p < .05$.

the TRAD students more external than the PAS dropouts. The IE average for PAS dropouts was comparable to the full sample average presented in the next section (Table 4). The TRAD dropouts were untypically external, but the sample size was much too small to draw any conclusions. The results suggest a need to explore the relationship between IE and dropout in other traditional introductory college courses.

All in all there is little evidence of differential dropout. The PAS and TRAD students who withdrew were similar in nearly all entry variables. The major difference was in the timing of their withdrawals; the PAS students dropped later in the term.

Group Differences on Entry Variables

This section assesses the equivalence of the PAS and TRAD students who remained in the course. In true experimental designs, the subjects are randomly selected and assigned to groups. Group equivalence is frequently assumed rather than tested because of the randomness in sampling. In the present study, the instructional methods were randomly assigned to the section, but strictly speaking, student membership in the sections was not randomly determined. The students did not know

that one section would be taught with nontraditional procedures when they signed up, but they were aware of the differing lecture schedules (two versus three lectures a week). Students with certain characteristics may have consistently selected one section over the other creating nonequivalent groups.

The subjects in this comparison were the 144 PAS and 156 TRAD students who took the final exam. Fourteen entry variables were originally selected for comparison, and several other variables were included later to clarify an issue. The sections appeared to be comparable on the two dichotomous variables, GENDER and MAJOR, and chi-square tests confirmed this (Table 5).

For the continous entry variables, regression analysis of variance was used with two levels of SECT by two levels of GENDER.[1] This design provided a comparison between genders as well as between sections. The F values for SECT, GENDER, and SECT by GENDER effects are presented in Tables 6 and 7, and for most variables, none of the effects were statistically significant.

---

[1]An assumption for analysis of variance is homogeneity of variance among the four subgroups (PAS males, PAS females, TRAD males, and TRAD females). This assumption was tested with Cochran's C Test. If there were indications of dissimilar variances, then the more lengthy but precise Bartlett's X^2 test was used. See Winer (1971) for a full description of these tests.

Table 5. Chi-square tests for relationship of SECT with GENDER and MAJOR for students who took the final

	PAS		TRAD		χ^{2}[a]
	obs.	exp.	obs.	exp.	
GENDER					
male	36	39.8	47	43.2	0.74
female	108	104.2	109	112.8	
MAJOR					
nonscience	99	101.8	113	110.2	0.33
science	45	42.2	43	45.8	

[a] χ^{2} calculated using Yate's correction, df = 1, $p < .05$ for both tests.

Table 6. Means, variances, n's, and summary of analysis of variance results on student entry variables

		PAS	TRAD	PAS		TRAD		F value[a, b]		
				male	female	male	female	SECT	GEND	S x G
ACT	$\bar{x}$	23.7	22.8	24.6	23.5	23.6	22.5	2.18	2.82	0.01
	var	19.2	19.5	16.5	20.0	13.7	21.6			
	n	107	126	27	80	36	90			
MSAT	$\bar{x}$	43.8	41.9	41.9	44.4	41.3	42.1	0.64	0.89	0.20
	var	161	139	173	158	148	136			
	n	109	131	27	82	40	91			
HSSCI	$\bar{x}$	4.65	4.51	4.81	4.60	4.50	4.51	0.60	0.15	0.20
	var	3.71	3.20	3.45	3.83	3.38	3.15			
	n	129	145	32	97	42	103			
RAM	$\bar{x}$	1.83	1.49	2.06	1.75	1.76	1.40	0.07	0.08	0.01
	var	82.2	64.0	82.1	83.0	40.5	72.2			
	n	138	138	33	105	34	104			
CONFID	$\bar{x}$	14.4	13.9	14.3	14.4	15.6	13.3	0.02	1.56	1.93
	var	38.8	43.3	35.3	40.3	31.9	46.1			
	n	139	139	33	106	35	104			
PLECT	$\bar{x}$	-0.87	-0.15	-0.97	-0.84	-0.40	-0.07	1.63	0.19	0.04
	var	13.2	14.7	12.7	13.4	18.3	13.7			
	n	138	140	32	106	35	105			
IE	$\bar{x}$	10.5	11.1	9.41	10.8	10.8	11.2	2.54	2.48	0.67
	var	19.0	14.5	19.1	18.7	14.2	14.7			
	n	137	135	33	104	35	100			

[a]For all tests for homogeneity of variance, $p > .05$.

[b]F's from 2 x 2 (SECT by GENDER) regression analysis of variance; $p > .05$ for all F tests.

Table 7. Means, variances, n's, and summary of analysis of variance results for student entry variables

				PAS		TRAD		F value[a, b]		
		PAS	TRAD	male	female	male	female	SECT	GEND	SxG
GPA	$\bar{x}$	2.71	2.60	2.51	2.77	2.45	2.67	1.36	8.23*	0.04
	var	.427	.415	.494	.393	.497	.367			
	n	139	153	34	105	47	106			
TOTHR	$\bar{x}$	66.2	65.1	88.4	58.8	72.2	62.0	1.40	13.0*	3.08
	var	1971	1803	2575	1571	1897	1748			
	n	144	156	36	108	47	109			
HSR	$\bar{x}$	21.8	24.1	31.6	18.5	32.5	20.6	0.41	28.5*	0.08
	var	360	286	604	239	376	212			
	n	127	145	32	95	42	103			
LGHSR[c]	$\bar{x}$	1.14	1.23	1.31	1.08	1.38	1.17	1.77	13.5*	0.03
	var	0.22	0.18	0.23	0.20	0.18	0.17			
	n	127	145	32	95	42	103			

[a]For all tests for homogeneity of variance, $p > .05$, except HSR (Bartlett's $\chi^2 = 211.0$, df = 3, $p < .01$).

[b]F's from 2 x 2 (SECT by GENDER) regression analysis of variance.

[c]LGHSR = $\log_{10}$ (HSR): transformation done to effect homogeneity of variance.

*$p < .01$.

GENDER differences were significant for GPA and TOTHR. On the average, the women exceeded the men by 0.2 grade points and had taken fewer college credit hours. The tests for homogeneity of variance for high school rank, HSR, revealed dissimilar variances. This was corrected by taking the logarithm of HSR to produce a new variable named LGHSR. Statistics for both HSR and LGHSR are given in Table 7. As with GPA, there were significant GENDER differences with women ranking higher in their respective high school classes than men.

Up to this point, twelve of the originally selected entry variables were tested for SECT difference, and no significant differences between the PAS and TRAD sections were found. Table 8 presents data on the two remaining variables, TAQ (test anxiety) and PALTT (preference for alternative testing procedures), which did differ significantly between sections. For TAQ there were significant GENDER and SECT main effects, and the S x G interaction approached significance. An inspection of Table 8 reveals that while the PAS female, TRAD male, and TRAD female TAQ means clustered in a tight range (73 to 75), the PAS male group was notably lower in TAQ (58), so low in fact, to produce the large F values. For PALTT only a significant SECT difference maintained; both

Table 8. Means, variances, n's, summary of analysis of variance results, and comparison of Spring 1977 and Winter 1977 on student entry variables

		PAS	TRAD	PAS		TRAD		F value[a, b]		
				male	female	male	female	SECT	GEND	S x G
TAQ	$\bar{x}$	69.7	74.9	58.3	73.3	73.3	75.4	6.16**	6.16**	3.45*
	var	736	510	707	698	674	461			
	n	139	139	33	106	34	105			
	($w\bar{x}$)[c]		(72.8)			(69.9)	(73.6)			
PALTT	$\bar{x}$	3.09	0.40	2.79	3.18	1.09	0.17	24.2**	0.30	1.86
	var	10.5	13.0	11.3	10.4	10.3	13.9			
	n	138	140	33	105	35	105			
	($w\bar{x}$)		(0.72)			(1.41)	(0.53)			
IEACAD	$\bar{x}$	1.58	1.96	1.39	1.63	1.80	2.01	6.00**	1.96	0.01
	var	1.46	1.15	1.37	1.49	1.22	1.13			
	n	139	136	33	106	35	101			
	($w\bar{x}$)		(1.86)			(1.72)	(1.89)			
IELUCK	$\bar{x}$	1.80	1.87	1.65	1.85	1.92	1.85	0.50	0.11	0.47
	var	1.35	1.41	1.37	1.35	1.56	1.37			
	n	137	138	33	104	35	103			
	($w\bar{x}$)		(1.81)			(2.21)	(1.71)			

[a]For all tests for homogeneity of variance, $p > .05$.

[b]F's from 2 x 2 (SECT by GENDER) regression analysis of variance.

[c]The means in parentheses ($w\bar{x}$) were from Winter Quarter, 1977. They were based on approximately 60 males and 235 females from two sections of traditionally taught Zoology 155.

*$p = .06$.

**$p < .01$.

sexes in PAS showed greater preference for alternative type testing procedures than the TRAD students.

The remainder of this section on initial group differences argues that in spite of SECT differences on TAQ and PALTT, PAS and TRAD were not different on entry variables before the beginning of Spring Quarter. It is hypothesized that the observed SECT differences in TAQ and PALTT were the result of the implementation of the Phase Achievement System, particularly the instructions to the PAS students, and not due to group differences prior to the quarter. Strictly speaking, the variables based on the pre-questionnaire, including TAQ and PALTT, were not true entry variables in the same sense as archival data such as ACT and GPA. The pre-questionnaire (Pre-Q) was administered during the third lecture period of the quarter. During parts of the first two meetings, the instructor and the developer of the Phase Achievement System introduced and explained the system to the PAS students. Because of the temporal sequence, the explanation of PAS procedures may have affected student response to the TAQ and PALTT items. Indirect evidence from two sources is presented to support this hypothesis.

An examination of the nature of the variables which did and did not show SECT differences provides evidence

on the tenability of the hypothesis. The PAS and TRAD sections did not differ significantly on any entry variables based on data existing prior to the quarter; that is, GENDER, MAJOR, ACT, MSAT, HSSCI, GPA, TOTHR, and HSR. Except for TAQ and PALTT, the other Pre-Q variables did not show SECT differences (RAM, CONFID, PLECT and IE). TAQ and PALTT consisted of items about testing, but the other Pre-Q variables did not except for a few items which were included in the academic factor of IE (IEACAD, see Chapter III for a description of the IE factors). If the hypothesis were true, then the PAS students should have scored more internally on the IEACAD scale than the TRAD students because of the nature of the IEACAD items while no differences would be expected for the IELUCK variable. The results presented in Table 6 support the hypothesis. Results from analysis of variance of other Pre-Q items also confirmed the hypothesis:

1.) The item on preference for study guide (Item 8) was not expected to differ between sections because explanations on the guide were similar for both sections. This was upheld (SECT $F = 0.10$, $df = 1, 273$; $p > .05$).

2.) SECT differences on Item 11 (preference for frequency of testing) were expected and found with PAS showing greater preference (SECT $F = 7.18$; $df = 1, 275$; $p < .01$).

3.) SECT differences on Item 12 (preference for normative grading) were expected and found with TRAD showing greater preference (SECT F =

3.85; df = 1, 275; $p < .05$).

The second source of support for the hypothesis lies in a comparison of PAS and TRAD means to Winter Quarter means (from two sections of traditional Zoology taught during the preceding Winter Quarter, 1977). The Winter means ($w\bar{x}$'s) in Table 7 support the premise that the typical TAQ score is in the low 70's and that the PAS male group was exceptionally low. It is not known why only the males reacted to the PAS instructions. Comparisons with Winter means are also provided for PALTT, IEACAD, and IELUCK. The Winter means were closer to the TRAD means than the PAS means for PALTT and IEACAD, indicating that the PAS means were deviant.

A sizeable majority of the variables did not show SECT differences. Indirect, but cogent evidence, was presented that supported the hypothesis that those few variables which did differ between sections were influenced by explanation of the PAS procedures to the students in the PAS section. The data indicate that by the third class meeting, the time of the Pre-Q, treatment effects were already present. That is, the PAS treatment produced a positive attitude toward alternative type testing and a more internal, academic locus of control for both sexes and reduced test anxiety for the males. It is not known

how lasting the effects were. The SECT differences observed are considered to be part of the PAS treatment. Therefore, it is concluded that there were no substantial group differences prior to the quarter and that the comparisons of the TRAD and PAS sections on outcomes presented in the next section are unhindered by initial group differences.

Group Differences on Outcomes and Study Variables

The ultimate goal of this chapter is to compare PAS and TRAD students on course outcomes. The two preceding studies were done to insure a valid test of SECT differences. Before the results are presented, the methodology particular to this section is explained.

The groups were compared using regression analysis of covariance with two levels of SECT and two levels of GENDER. Covariance analysis was selected to increase statistical precision, not as in many studies to adjust for nonequivalent groups. The covariate was chosen mainly with the SCORE outcomes in mind, and GPA provided an excellent covariate for the following reasons: it is highly related to academic performance and, hence, affords excellent precision in the analysis of SCORE; GPA was missing for only eight subjects; and it met the statistical

assumptions for a covariate.[1] For the sake of consistency, covariance was used with the attitudinal outcomes and study variables.

Special attention was given to the subjects included in the analysis of outcomes. In the PAS section there were 28 students who did not pass all nine phases before taking the comprehensive final. Ten of these students took incompletes, and the remainder either received a failing grade or took make-up phase tests after the final exam but before grades were forwarded to the Registrar's Office. Of these students, 14 passed 8 phases, 9 passed 7, 2 passed 6, 1 passed 5, 1 passed 2, and 1 student passed 1 phase. Consequently, a problem arose in deciding which of these students to include in the comparison between the PAS and TRAD sections, particularly among those who passed only a few phases. They should not be excluded because of poor performance because this would bias the results in favor of PAS. On the other hand, they should not be included if they were not "legitimate" PAS students, students who did not actively

---

[1]Analysis of covariance assumes the covariate is unaffected by the treatment, and with GPA being an archival variable, this assumption was met. Another requirement is that the main effects do not interact with the covariate. It was found that SECT and GENDER did not interact with GPA as they related to SCORE. Homogeneity of variance was also tested as described in the previous section.

participate and take advantage of the PAS system. It was decided to judge the legitimacy of the students on the basis of the total number of phase tests attempted before the final exam. The average number of tests attempted for the 28 students was 16.4 with a range from 6 to 34. Of prime interest was the number of tests attempted by the two students who passed only 1 and 2 tests. They attempted 12 and 13 tests respectively, indicating that they were relatively active. The subjects then included all students taking the final who had a valid GPA value.

Cognitive and attitudinal outcomes

Table 9 presents descriptive data and results of analysis of covariance for the four outcomes. The F values refer to differences in group means adjusted for GPA. The SECT difference for SCORE was not statistically significant nor did it approach significance. The PAS students, on the average, scored about the same as the TRAD students on the comprehensive final. The sexes did differ significantly with the men, as a group, exceeding the women by three to four points after adjusting for GPA. The interaction was not significant[1], and the large

[1]The GENDER by SECT interaction is actually an attribute by treatment interaction with sex as the attribute. The nonsignificant interaction indicates that PAS and TRAD instruction were not differentially effective (in terms of SCORE) for either sex.

Table 9. Means, means adjusted for GPA, variances, n's, and results from analysis of covariance on outcomes

		PAS	TRAD	PAS		TRAD		F value[a, b]	
				male	female	male	female		
SCORE	$\bar{x}$	58.0	55.9	58.6	57.8	56.3	55.7	SECT F=	0.95
	adj $\bar{x}$	57.4	56.4	60.2	56.4	58.6	55.5	GEND F=	6.83**
	var	131.4	163.1	106.2	140.5	174.6	159.5	S x G F=	0.06
	n	139	153	34	105	47	106	GPA	F=149.76**
F1	$\bar{x}$	4.31	4.07	4.01	4.40	3.95	4.11	SECT F=	1.52
	adj $\bar{x}$	4.33	4.08	4.05	4.42	3.99	4.11	GEND F=	1.18
	var	1.79	2.34	1.90	1.75	2.15	2.43	S x G F=	0.26
	n	127	131	29	98	35	96	GPA F=	2.32
F2	$\bar{x}$	3.29	3.79	2.84	3.42	3.63	3.85	SECT F=	8.36**
	adj $\bar{x}$	3.29	3.79	2.78	3.45	3.58	3.86	GEND F=	5.15*
	var	1.95	2.07	1.07	2.15	1.67	2.22	S x G F=	0.93
	n	128	131	30	98	35	96	GPA F=	3.84*
INSTR	$\bar{x}$	0.33	0.52	0.66	0.23	0.64	0.47	SECT F=	0.32
	adj $\bar{x}$	0.33	0.52	0.67	0.23	0.65	0.47	GEND F=	2.33
	var	1.95	1.80	1.60	2.03	1.53	1.91	S x G F=	0.44
	n	128	127	30	98	33	94	GPA F=	0.03

[a]For all tests for homogeneity of variance, $p > .05$.

[b]F's from 2 x 2 (SECT by GENDER) regression analysis of covariance; GPA was the covariate.

*$p \leq .05$.

**$p \leq .01$.

F value for GPA simple means that GPA was significantly related to SCORE and that the use of the covariate increased the precision of the analysis.

Student attitudes about instruction were compared next. The F1 variable reflected perceived increase of interest in zoology and general evaluation of the course. The theoretical range on F1 was -0.1 to 7.9, and a student responding halfway between the agree-disagree poles on all nine of the F1 items would have scored 3.9. All subgroups were somewhat above the neutral point but none dramatically so.

For the variable F2, the higher the scores the more negative the students were about tests and the more difficult they perceived the course standards. The variable also appeared to have measured student opinion about the fairness of the procedures. The theoretical range was -0.3 to 7.8 with 3.8 as the neutral point. The TRAD mean, sexes combined, fell at the neutral point and the PAS mean was significantly below. As a group the PAS students perceived the course as more equitable than the TRAD students. There was also a significant GENDER difference with males more positive than females. These differences in attitude did not seem to color the students' rating of the instructor. There were neither

SECT nor GENDER differences for INSTR.

Several of the post-questionnaire items which did not load highly on the F1 or F2 factors were analyzed separately. The four items presented in Table 10 had ranges of 1 to 9 and a neutral point of 5. Items 20 and 21, degree of self-determination of course grade and use of test results to adjust study patterns, were not strictly outcomes. They were analyzed more to study the degree to which the PAS students used the Phase Achievement System than to compare them with TRAD. The TRAD means were close to the neutral point while the PAS students were significantly higher, indicating that at least on the basis of student reports, PAS was successfully implemented.

As discussed in the preceding section on initial group differences, the PAS group showed greater preference for alternative testing procedures. The evidence suggested that this was due to instructions about the Phase Achievement System to the PAS students. The results from two post-questionnaire items (Items 33 and 34 in Table 9) indicated that the preference was still apparent at the end of the quarter. The PAS section was significantly more positive toward the self-paced, mastery (versus normative) testing components than the TRAD section.

Table 10. Means, means adjusted for GPA, variances, n's, and results from analysis of covariance on outcomes

	PAS	TRAD	PAS		TRAD		F value[a, b]
			male	female	male	female	
ITEM 33 $\bar{x}$	6.10	4.95	6.46	6.00	5.34	4.80	SECT F= 9.13**
adj $\bar{x}$			6.40	6.02	5.28	4.81	GEND F= 1.20
var	6.79	7.20	4.41	7.47	7.41	7.13	S x G F= 0.13
n	126	131	28	98	35	96	GPA F= 1.66
ITEM 34 $\bar{x}$	5.66	7.32	5.10	5.83	6.91	7.47	SECT F=28.13**
adj $\bar{x}$			5.21	5.84	6.88	7.47	GEND F= 4.89*
var	6.72	3.47	6.60	6.70	3.67	3.35	S x G F= 0.08
n	127	131	29	98	35	96	GPA F= 0.88
ITEM 20 $\bar{x}$	6.25	4.88	6.30	6.23	4.77	4.85	SECT F=27.33**
adj $\bar{x}$			6.15	6.24	4.78	4.86	GEND F= 0.01
var	4.09	3.33	3.53	4.31	3.42	3.33	S x G F= 0.07
n	128	131	30	98	35	96	GPA F= 0.01
ITEM 21 $\bar{x}$	5.96	5.37	5.70	6.04	4.94	5.53	SECT F= 5.76**
adj $\bar{x}$							GEND F= 2.16
var	3.50	3.13	2.83	3.71	3.53	2.93	S x G F= 0.27
n	128	131	30	98	35	96	

[a]For all tests for homogeneity of variance, $p > .05$.

[b]F's from 2 x 2 (SECT by GENDER) analysis of covariance; GPA was the covariate.

*$p < .05$.

**$p < .01$.

Study pattern variables

The number of study hours spent outside of lecture, HRSPERQ, showed unequal subgroup variances. The log transformation of HRSPERQ to LGHRPQ resulted in more homogeneous variances and also produced a more symmetric distribution as illustrated in Figure 1. The more accurate F values in Table 11 are the ones listed with LGHRPQ, however, the significant SECT and GENDER effects are best interpreted by studying the adjusted means for HRPERQ. It can be seen the subgroups averaged 32 to 39 hours of reported outside study time except the PAS females who averaged 54 hours for the quarter. Eight of the nine students who responded "more than 10 hours" of study time per week--this was translated to 120 hours for the entire quarter--were PAS females. The bunching of PAS females at the high end of HRSPERQ is illustrated in Figure 1. The interpretation is that the PAS students and the females, on the average, reported spending significantly more study time than the TRAD students and the males, but both of these differences were due primarily to the PAS females. The nonsignificant t-test, comparing the TRAD males with the PAS males on HRSPERQ, supports this interpretation ($t = 0.71$, $df = 63$; $p > .05$).

One of the items from the F2 factor, Item 24, was

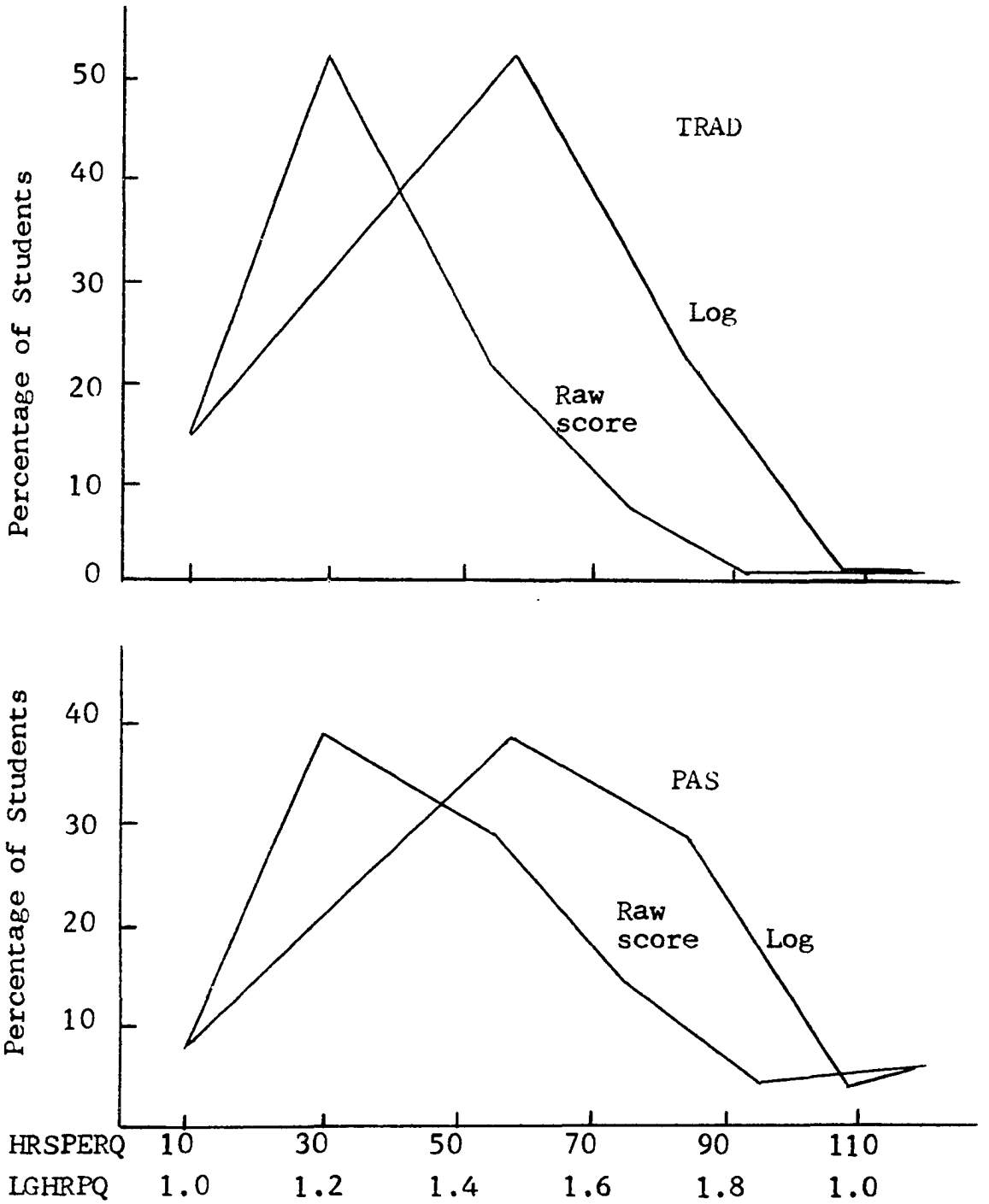


Figure 1. Frequency distributions of HRSPERQ and LGHRPQ for PAS and TRAD students

Table 11. Means, means adjusted for GPA, variances, n's, and results from analysis of covariance on study patterns

		PAS	TRAD	PAS		TRAD		F value[a, b]
				male	female	male	female	
HRSPERQ	$\bar{x}$	50.0	37.2	37.4	54.0	32.2	38.9	SECT F= 9.21**
	adj $\bar{x}$	49.9	37.1	35.1	54.5	31.7	39.0	GEND F=13.24**
	var	759	403	371	819	376	405	S x G F= 2.27
	n	130	130	31	99	34	96	GPA F= 2.15
LGHRPQ[c]	$\bar{x}$	1.62	1.50	1.51	1.67	1.43	1.53	SECT F= 9.07**
	adj $\bar{x}$	1.63	1.50	1.49	1.67	1.43	1.53	GEND F=11.88**
	var	0.07	0.07	0.06	0.07	0.08	0.06	S x G F= 0.06
	n	130	130	31	99	34	96	GPA F= 0.38
TEXT%	$\bar{x}$	4.03	2.92	4.10	4.01	3.00	2.89	SECT F=30.39**
	adj $\bar{x}$	4.02	2.91	4.09	4.00	3.01	2.88	GEND F= 0.18
	var	1.89	2.06	1.62	1.99	2.12	2.06	S x G F= 0.01
	n	128	131	31	97	35	96	GPA F= 0.15
SGUID%	$\bar{x}$	3.85	4.08	3.71	3.89	3.37	4.34	SECT F= 0.17
	adj $\bar{x}$	3.82	4.12	3.77	3.84	3.52	4.34	GEND F= 7.07**
	var	1.77	1.54	1.81	1.77	1.95	1.15	S x G F= 5.56*
	n	130	131	31	99	35	96	GPA F=13.03**

[a]For all tests for homogeneity of variance, $p > .05$ except HRSPERQ (Bartlett's $\chi^2 = 13.86$, df = 3, $p < .01$).

[b]F's from 2 x 2 (SECT by GENDER) regression analysis of covariance; GPA was the covariate.

[c]LGHRSPQ = LOG_{10} (HRSPERQ); transformation done to effect homogeniety of variance.

*$p < .05$.

**$p < .01$.

analyzed separately because it was particularly relevant to HRSPERQ. The item asked the students to agree-disagree with the statement, "Compared to other courses I took this quarter, I spent too much time on this course for the credit assigned." The logarithm of the item, LGITEM 24, was calculated to produce homogeneous variances. The results are presented in Table 12. Although the PAS females reported more outside study hours (HRSPERQ) than the other subgroups, they did not perceive themselves as spending too much time when compared to the subgroups.

TEXT% and SGUID% measured the reported percentage of completion of text assignments and study guide. These process variables provided some clues about where the students' time was spent, but the evidence was indirect because they did not refer to the actual number of study hours. The PAS group reported reading a significantly greater amount of the text than the TRAD group. The PAS adjusted mean was about 4 versus about 3 for the TRAD section; these values translate to 61 - 80% versus 41 - 60% completion of the assignment. For SGUID% the sections were about equal with both reporting, on the average, 61 - 80% mastery of the study guide. An inspection of the means and the SECT by GENDER F value reveals that the significant GENDER difference is due

Table 12. Means, means adjusted for GPA, variances, n's, and results from analysis of covariance on outcomes

		PAS	TRAD	PAS		TRAD		F value[a, b]	
				male	female	male	female		
LGITEM 24[c]	$\bar{x}$	0.62	0.57	0.58	0.64	0.59	0.56	SECT	F=0.85
	adj $\bar{x}$	0.62	0.57	0.57	0.64	0.58	0.57	GEND	F=0.71
	var	0.07	0.05	0.06	0.07	0.05	0.05	S x G	F=1.79
	n	126	131	30	96	35	96	GPA	F=5.27*
ITEM 14	$\bar{x}$	5.83	5.59	5.50	5.93	5.43	5.65	SECT	F=0.27
	adj $\bar{x}$	5.82	5.57	5.57	5.89	5.50	5.60	GEND	F=0.48
	var	4.92	4.58	5.50	4.75	4.02	4.82	S x G	F=0.09
	n	128	131	30	98	35	96	GPA	F=4.34*
ITEM 28	$\bar{x}$	6.48	7.20	6.57	6.46	6.86	7.33	SECT	F=4.09*
	adj $\bar{x}$	6.47	7.22	6.65	6.42	6.94	7.32	GEND	F=0.06
	var	4.90	3.57	3.91	5.24	3.36	3.63	S x G	F=1.07
	n	128	130	30	98	35	95	GPA	F=5.24*

[a]For all tests for homogeneity of variance, $p > .05$.

[b]F's from 2 x 2 (SECT by GENDER) regression analysis of covariance; GPA was the covariate.

[c]LGITEM 24 = LOG_{10} (Item 24); transformation done to effect homogeneity of variance.

*$p \leq .05$.

mostly to the high percentage of completion reported by the TRAD females.

The high number of outside study hours (HRSPERQ) reported by the females does not seem to be accounted for by special effort on the text or study guide because the subgroup pattern of means for TEXT% or SGUID% did parallel the HRSPERQ pattern. Possibly the PAS students included the time they spent on taking phase tests in their estimates of HRSPERQ and the PAS females took more tests than the males. A t-test on the total number of tests attempted, however, showed no sex differences in the PAS section ($t = 0.40$, $df = 137$, $p > .05$). The only other resource that might have accounted for HRSPERQ was time spent with the lecture notes, but there were no measures of this variable.

The students' evaluation of the quality of the textbook and study guide (Items 14 and 28 in Table 11) did not appear to relate strongly to the reported usage of these resources. While the PAS students reported greater completion of the text assignments, they did not rate the text higher than the TRAD group. The sections did not differ on SGUID%, but the TRAD students as a group felt the study guide was more useful. Overall, the students reported greater completion of the

content than the assigned readings (grand mean TEXT% = 3.47, or about 60%; grand mean SGUID% = 3.97, or about 70%; paired t = 4.28, df = 258, $p < .01$ level).

CHAPTER V. STUDY OF RELATIONSHIPS

The second purpose of this research is to study achievement, attitudes, and study behavior in introductory zoology by examining the relationships among variables in the overall group. The methods are discussed and then the findings are presented according to the outline below:

1. Correlations among entry, process, and outcome variables. By studying correlations within these groups of variables it is possible to more fully understand the meaning of the variables.
2. Prediction of SCORE. Major predictors of SCORE are identified, and the contribution of questionnaire variables is examined.
3. Prediction of attitudinal outcomes.
4. Prediction of process variables.

Methods

The basic statistic in this chapter and the next is the simple Pearson correlation coefficient (zero-order correlation) which is a measure of the linear relationship between two variables. Such a correlation can be calculated for any two variables for descriptive

purposes regardless of the distributions. Ideally, however, the two variables should be similarly distributed because this condition allows a maximum correlation. In other words, if the distributions are markedly different in shape, the calculated correlation will underestimate the true relationship. Furthermore, if statistical tests of significance are made, the two variables should display a reasonably normal, bivariate distribution. If the two variables are reasonably normally distributed, then in most cases, they satisfy the assumption of a normal, bivariate relationship.

The overall (sexes and sections combined) distributions of each variable were visually examined. If a distribution deviated considerably from a symmetrical, bell-shaped pattern, then the variable was transformed and the resulting distribution was reexamined. The decision to transform a variable was made on the basis of the shape of a distribution, not its correlation with an outcome. The distributions of TOTHRS, HSR, IELUCK, IEACAD, and HRSPERQ were positively skewed (bunching at the left side). A logarithmic transformation produced more symmetrical distribution, and consequently, the transformed variables, LGTOTH, LGHSR, LGIELK, LGIEAC, and LGHRPQ were substituted for the original variables. The other

distributions were approximately bell-shaped except TEXT% and SGUID%. These were very positively skewed, and none of the possible transformations produced more normal shapes. Because of the skewed distributions, the correlations for TEXT% and SGUID% were minimal estimates.

The first step in the analysis was to generate a master table including correlations among all variables for the overall group (see Table 13). There were 17 entry variables, 3 process variables, and 4 outcomes. Sections of the master table were then used in considering the specific objectives. Missing data were handled with pair-wise deletion; that is, if a value were missing for a student on either variable, the student was eliminated only when calculating that particular correlation. Appendix B lists the number of cases for each correlation listed in Table 13.

All correlations were tested for significance using t-tests (Nie et al., 1975). The tests were two-tailed with no predictions about the signs of the correlations. Obtaining statistical significance means that a correlation differs significantly from zero and is a function of the magnitude of the correlation and the sample size. In this chapter some correlations of .12 are significant

Table 13. Pearson correlation matrix for overall group (sections and genders combined)[a]

	n max	LGTOTH	MAJOR	LGHSR	GPA	MSAT	ACT	HSSCI	CONFID	PLECT	PALTT	HS	FF
LGTOTH	300												
MAJOR	300	-12*											
LGHSR	272	-07	-12*										
GPA	292	19*	-06	-64*									
MSAT	240	18*	10	-53*	47*								
ACT	233	16*	14*	-56*	52*	73*							
HSSCI	274	-16*	20*	-22*	11	23*	33*						
CONFID	278	-16*	35*	-12*	03	22*	24*	39*					
PLECT	278	05	-09	05	04	01	00	02	-06				
PALTT	278	12*	-07	-03	-07	-05	-01	-02	-06	-17*			
HS	277	14*	01	03	-09	03	10	00	17*	-07	08		
FF	278	-11	06	09	-11	-11	-12	-17*	-16*	12*	00	-13*	
RAM	276	16*	-03	-05	01	10	14*	13*	21*	-13*	05	72*	-78*
TAQ	278	-09	-05	22*	-26*	-36*	-40*	-11	-22*	-01	10	-14*	34*
IE	272	05	-07	03	-08	-11	-17*	-05	-23*	09	-03	-07	22*
LGIECK	275	03	-01	13*	-09	-18*	-17*	-04	-19*	13*	03	-09	14*
LGIEAC	275	-02	-05	09	-16*	-12	-19*	-08	-15*	07	-08	00	12*
LGHRPQ	268	-12	10	-02	01	-14*	-17*	-12	-01	-17*	05	01	05
TEXT%[b]	267	-02	00	04	-01	-01	01	-10	23*	-10	16*	11	-03
SGUID%[b]	269	-14*	-01	-19*	23*	12	09	-02	06	-01	-10	-02	11
F1	266	-10	32*	-11	09	08	12	19*	46*	02	-11	17*	-18*
F2	267	01	-24*	09	-12	-09	-25*	-25*	-36*	07	03	-15*	10
INSTR	263	05	-04	04	-01	-03	06	01	05	10	-04	06	-05
SCORE	300	03	20*	-44*	58*	32*	42*	25*	32*	01	-14*	-06	-05

[a] n's are listed in Appendix B. n max is the maximum n for a variable.

[b] Skewed distribution; correlations are minimal estimates.

* $p \leq .05$; two tailed test.

RAM	TAQ	IE	LGTELK	LGTEAC	LGHRPQ	TEXT%	SGUID%	F1	F2	INSTR
-32*										
-19*	33*									
-15*	17*	72*								
-08	32*	62*	36*							
-04	05	00	03	05						
09	-01	-09	-06	-11	37*					
-10	09	01	-01	04	25*	10				
23*	-11	-16*	-11	-07	22*	12	22*			
-16*	27*	21*	20*	19*	11	08	04	-44*		
08	-03	-15*	-11	-04	00	-06	-08	44*	-35*	
00	-24*	-04	-01	-13*	12*	08	27*	37*	-30*	06

in the overall group, but in smaller groups such as the male PAS or TRAD section, correlations as large as .34 are not significant. Another way to interpret correlations is to square the value; the resulting number, called the coefficient of determination, indicates the percentage of variance in one variable explained by or shared with the other variable. Curvilinearity between two variables, say X and Y, was tested by examining the contribution of X^2 in predicting Y after X was allowed to predict Y (see Kerlinger and Pedhazur, 1973).

Correlations Among Entry, Process, and Outcome Variables

Several patterns were present among the entry variables in the overall correlation table (Table 13). The intercorrelations of four of the cognitive entry variables, LGHSR, GPA, MSAT, and ACT, were particularly high. They ranged in absolute value from .47 to .73 with an average of .58. This aptitude cluster as a whole correlated only mildly with HSSCI, MAJOR, and CONFID, which themselves tended to form a group (r's = .20, .35, and .39). TAQ was the only personality measure to correlate consistently with the aptitude cluster; the greater the test anxiety, the lower the general aptitude,

particularly as measured by tests (ACT, MSAT). True to the mixed content of its items (interest in zoology and self-rated aptitude), CONFID was related to cognitive (ACT, MSAT) and noncognitive characteristics (RAM, TAQ, and IE).

The personality scales showed some clustering but not as apparent as with the aptitude measures. The subscale to full scale correlations (HS and FF to RAM and LGIELK and LGIEAC to IE) were very high because of item redundancy. The intercorrelations for RAM, IE, and TAQ were moderate (-.19, -.32, .33), but the subscale and TAQ variables definitely did not form a cohesive grouping (see Table 13).

Preference for instructional methods as measured by PLECT and PALTT had little to do with aptitude or personality. Their uniqueness is seen in the fact that they correlated higher with each other (-.17) than with any other variable.

The self-reported study variables included number of study hours per quarter (LGHRPQ) and the mastery percentage of the study guide (SGUID%) and text assignments (TEXT%). These last two were essentially unrelated (r = .10), indicating they were not typically

used together.[1] Lecture notes may have been used frequently in conjunction with the study guide, but unfortunately, this process variable was not measured. Number of study hours correlated .37 and .25 with SGUID% and TEXT%. The multiple correlation for SGUID% and TEXT% working together to predict LGHRPQ was .43, or in other words, 18% of LGHRPQ was accounted for jointly by these two. This low figure may have been due to several reasons: the unit of measurement for LGHRPQ was hours spent while the others were percentage of use; there may have been large measurement error because of self-reports; and study of lecture notes was included as a variable.

The correlations among outcomes are presented below:

	F1	F2	INSTR
F2	-.44		
INSTR	.44	-.35	
SCORE	.37	-.30	.06

A relatively strong cluster emerged among the attitudinal outcomes, F1, F2, and INSTR. The directions of the relationships were not surprising. Those who rated the general course high and reported an increase in interest in life sciences (F1) tended also to hold a more positive feeling toward the testing and grading procedures (F2)

[1]The low correlation was not due to disparate distributions because both were similarly skewed.

and rated the instructor higher (INSTR). The F1 and F2 correlation, while fairly high, supports the factor analysis finding that they are not simply poles of one continuum. Opinions about instruction (F1 and F2) were related to SCORE, but opinion about the instructor was independent of SCORE.

Prediction of SCORE

Table 14 presents the correlations between all entry and process variables and SCORE on the final exam for the overall group. The cognitive entry variables (GPA, LGHSR, ACT, MSAT, and HSSCI) were all moderately to strongly related to performance on the final exam. College GPA stands out as the best predictor, explaining about 34% of the variability of SCORE. The HSSCI correlation was not as high as the others, but this was due to the fact that it correlated highly in one section and low in the other. This difference in correlations is discussed in the next chapter.

All of the cognitive variables were linearly related to SCORE except ACT. The square of ACT significantly contributed to the prediction ($df = 1, 230$, $F = 5.51$, $p < .05$). Figure 2 shows the best fitting curve ($SCORE = 66.06 - 2.26\,[ACT] + .08\,[ACT]^2$), and for comparison, the

Table 14. Correlations between entry and process variables and SCORE in the overall group[a]

	n	simple r	part r[b]
GPA	292	58*	00
LGHSR	272	-44*	-09
ACT	233	42*	16*
MSAT	240	32*	05
HSSCI	274	25*	19*
CONFID	278	32*	30*
MAJOR	300	20*	23*
TAQ	278	-24*	-09
LGIEAC	275	-13*	-04
PALTT	278	-14*	-10
SGUID%[c]	269	27*	14*
LGHRPQ	268	12*	12*
TEXT%[c]	267	08	09
LGTOTH	300	03	-08
PLECT	278	01	-01
HS	277	-06	-02
FF	278	-05	01
RAM	276	00	-01
IE	272	-04	01
LGIELK	275	-01	04

[a]Sections and sexes combined, abstracted from Table 13.

[b]GPA partialed from predictor variable.

[c]Skewed distributions, correlations are minimal estimates.

*$p < .05$; two-tailed test.

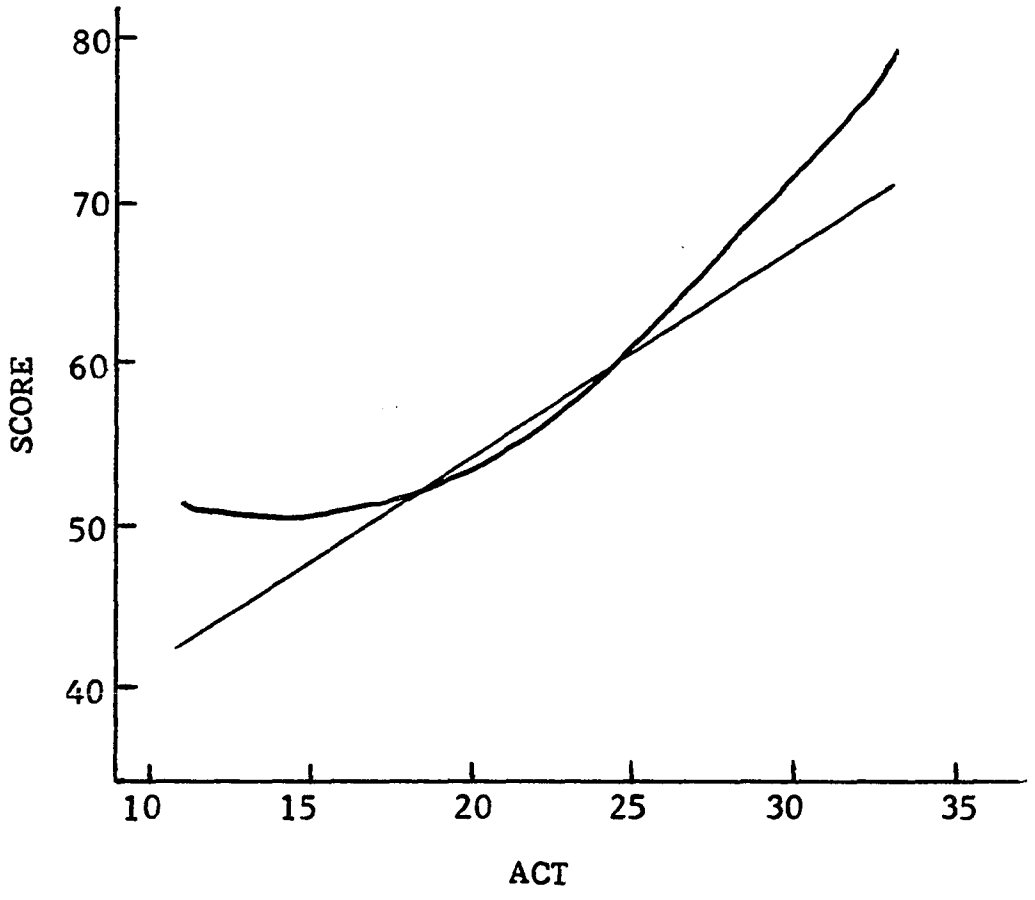


Figure 2. Best fit linear and curvilinear lines from regression of SCORE on ACT

best fitting straight line is also plotted (SCORE = 29.81 + 1.17 [ACT]). At the low end of ACT, the variable is not related to SCORE, but after ACT scores of about 20, the relationship is quite strong.

As a group, the preference variables (PALTT and PLECT), and the personality traits (TAQ, IE, LGIEAC, LGIELK, RAM, HS, and FF) were weakly related to SCORE. The strongest of this affective group was TAQ with the more anxious scoring lower on the final. Although several researchers have speculated and found nonlinear relationships between anxiety and achievement, nonlinearity was not found in the sample. All affective variables were tested for curvilinearity and none were statistically significant.

The correlation between the full scale IE and SCORE was essentially zero, and the scale was factored to identify subscales which might be more strongly related (see Chapter III). The correlations for the LGIEAC and LGIELK subscales, however, showed little improvement in the prediction of SCORE in the overall group.

Of the process variables, only SGUID% explained any sizable variance in SCORE (7%). The correlation between study hours, LGHRPQ, and SCORE was disappointingly low. One would expect higher than the observed correlation of .12, and therefore extra attention was paid to LGHRPQ.

The addition of the square term was nearly significant (df = 1, 265, F = 3.73, p = .06), but adding the term raised the correlation from .12 to only .17. The relationship was further investigated by examining interactions with entry variables. Because the number of study hours spent might correlate higher with SCORE for brighter students, several cognitive variables (GPA, LGHSR, ACT, MSAT) were tested for interaction with LGHRPQ.[1] None were significant nor did they approach significance, indicating that the LGHRPQ correlation was small for the high as well as the low aptitude students. In addition to the aptitude measures, TAQ, CONFID, and LGTOTH (total number of college credit hours) were tested, and of these, the LGTOTH by LGHRPQ interaction was close to significance (df = 1, 264, F = 3.11, p = .08).

This interaction trend can be illustrated and interpreted by dichotomizing total credit hours into freshmen and upperclassmen. For the freshmen (n = 124), the correlation between LGHRPQ and SCORE was .03 and linear. For upperclassmen, the correlation was considerably higher at .21, and the nonlinear term was significant (df = 1, 141, F = 5.60, $p < .05$) which boosted the correlation

---

[1]This was done by allowing the aptitude measure and LGHRPQ to predict SCORE and then testing the contribution of the interaction term for statistical significance. See Kerlinger and Pedhazur (1973).

from .21 to .28. Separate best fit lines are plotted in Figure 3 (freshmen: SCORE = 56.03 + 1.19 [LGHRPQ]; upperclassmen: SCORE = -10.75 + 84.11 [LGHRPQ] - 25.42 $[LGHRPQ]^2$). The number of study hours was related to SCORE only in certain circumstances--when students were upperclassmen--and then, only at low and average levels of study hours. In spite of the efforts, the relationship was still modest. This may have occurred because the study hours variable was self-reported and students had to estimate it at the end of the quarter.

As noted before, GPA was the best single predictor of SCORE. GPA was also strongly related to other predictors such as LGHSR, MSAT, ACT, and TAQ. For the purpose of parsimony, the contribution of other variables above and beyond GPA was investigated. The method used was to calculate a part correlation, which is a measure of the relationship between SCORE and another variable with the effect of GPA statistically removed from that variable.[1] The degree to which a part

---

[1]A part correlation, also called semipartial correlation, is conceptually obtained by regressing GPA on MSAT, computing a residual for each student (observed MSAT minus predicted MSAT from GPA) and calculating a simple correlation between the residuals and SCORE (see Nunnally, 1967).

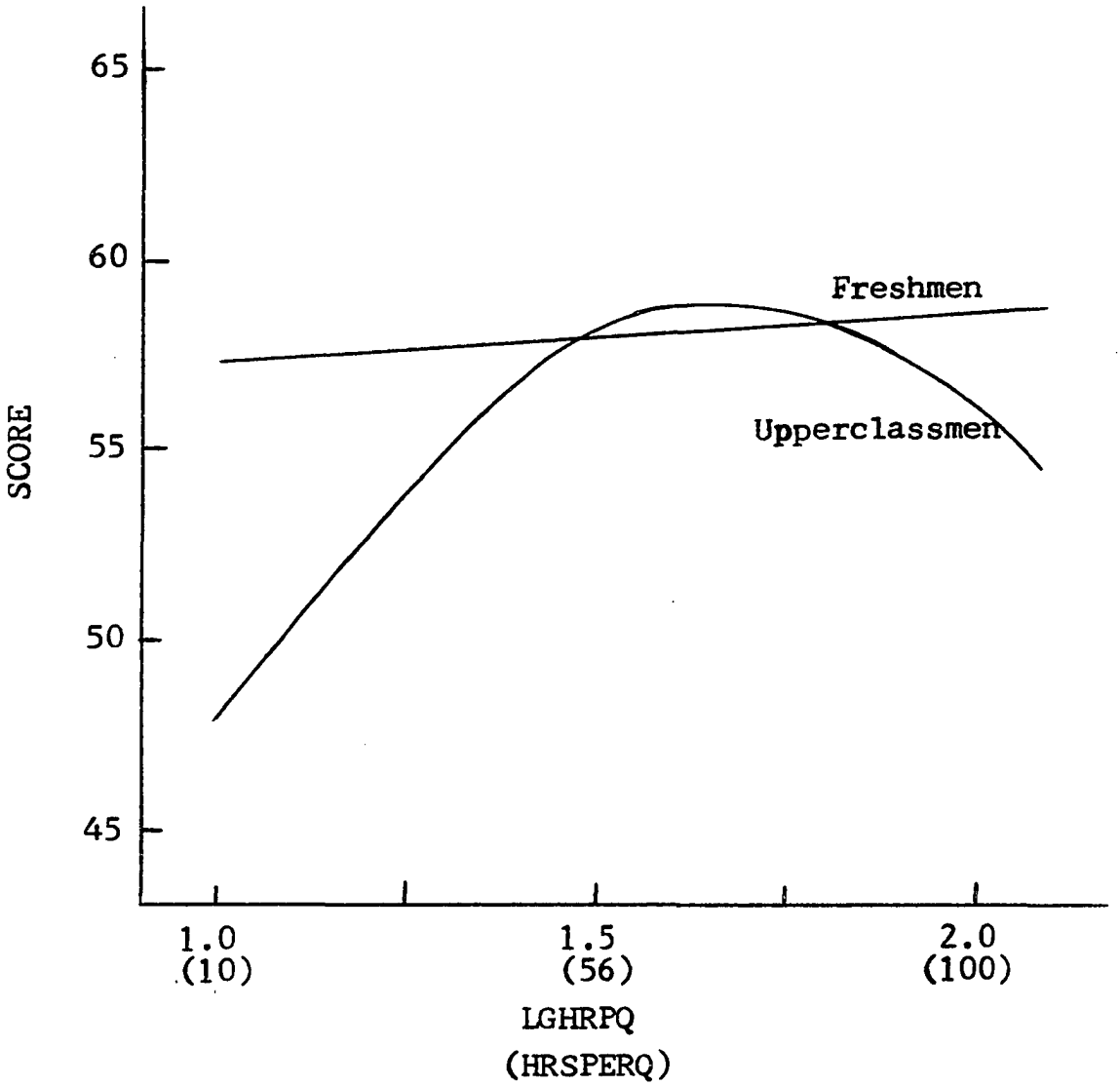


Figure 3. Best fit lines from regression of SCORE on LGHRPQ. HRSPERQ is included for ease of interpretation.

correlation is high is the degree to which a variable is related to SCORE above and beyond GPA.

Table 14 lists part correlations for all entry and process variables. Six of the 11 significant predictors maintained a statistically significant part correlation. High school rank (LGHSR) dropped from -.44 to -.09 because it was highly related to GPA, while CONFID decreased little because it was only modestly related to GPA. The increase in part correlation for MAJOR was due to the slight negative correlation between MAJOR and GPA (this example of cooperative suppression is explained in Cohen and Cohen, 1975). The part correlation procedures identifies the best of all possible pairs of predictors--in this sample, they were GPA and CONFID.

One of the subgoals of the first research purpose was to investigate the utility or necessity of the pre- and post-questionnaire variables in the prediction of SCORE. The administration of the questionnaires took about one and a half hours of student time away from the lectures plus many hours of staff time arranging make-ups. If questionnaire data contributed little to the prediction of SCORE beyond the more readily available archival data, then considerable time could be saved by eliminating the questionnaires. The part correlation approach is

appropriate for this objective, and a large part of the question can be answered from data results in Table 14. The column of part correlations show that CONFID, SGUID%, and LGHRPQ were the only questionnaire variables that maintained significant predictive power after the archival GPA was allowed to predict SCORE. The procedure then was to examine part correlations of the three variables beyond all archival variables.[1] The part correlation for CONFID was still fairly high (.22) and explained about 5% of the variance in SCORE above and beyond the archival variables. The part correlations for SGUID% and LGHRPQ were .12 and .13, statistically significant but low. For the most part, questionnaire data beyond CONFID was useless in predicting SCORE. There are, of course, other objectives besides predicting SCORE.

Prediction of Attitudinal Outcomes

Only the significant predictors of F1, F2, and INSTR in the overall group are listed in Table 15. Student rating

---

[1]The use of all archival variables (GPA, LGHSR, HSSCI, MAJOR, LGTOTH, ACT AND MSAT) would have reduced the number of cases from about 300 to 200, but by eliminating ACT and MSAT the sample size decreased to only 250, ACT and MSAT therefore were dropped from the analysis, leaving five archival variables.

Table 15. Correlations for significant predictors of attitudes in the overall group[a]

	F1		F2		INSTR	
	r	n	r	n	r	n
CONFID	46*	255	-36*	256	05	253
MAJOR	32*	266	-24*	267	-04	263
HSSCI	19*	244	-25*	245	01	241
ACT	12	212	-25*	213	01	210
RAM	23*	252	-16*	253	08	250
HS	17*	253	-15*	254	06	251
FF	-18*	254	10	255	05	252
IE	-16*	249	21*	250	-15*	248
LGIEAC	-07	252	19*	253	04	250
LGIELK	-11	252	20*	253	-11	251
TAQ	-11	254	27*	255	-03	252
SGUID%[b]	22*	266	04	267	-08	263
LGHRPQ	22*	265	11	266	00	262

[a]Sexes and sections combined.

[b]Skewed distribution, correlations are minimal estimates.

*$p < .05$; two-tailed test.

of the instructor, INSTR, was unpredictable. Of all the variables, only one was significant, IE, and that bore only a mild relationship. The remaining comments are confined to F1, a measure of overall satisfaction with course and interest in the subject matter, and to F2, a measure of feelings of unfairness in the testing and grading procedures.

The various correlations for F1 and F2 tended to be of similar magnitude and opposite in direction. CONFID, a measure of interest and expectancy of success in zoology, was the most consistent and strongest predictor of attitudes; it correlated higher with F1 and F2 than with SCORE. There was a trend for students who were scientifically oriented (CONFID, MAJOR, HSSCI) to be more positive toward the course while general aptitude (GPA, LGHSR, MSAT, ACT) was largely unrelated. The correlations for the preference variable (PLECT and PALTT) were near zero while the personality scales (RAM, IE and TAQ) were modestly related to F1 and F2. TAQ was the second strongest correlate of F2; the correlation was only .27, indicating that F2 was not a measure of just test anxiety. The process variables related only to F1 with those students showing more effort, giving the course a higher overall rating.

Prediction of Study Variables

Table 16 includes entry variables which correlated significantly with any of the study variables. The data reveal that study patterns were largely unpredictable in the overall group. Those few correlations which were significant were quite modest. This may have been due, in part, to the skewed distributions of the study variables. One interesting result is the lack of agreement among the general aptitude measures. Those who studied less (LGHRPQ) tended to be higher on ACT and MSAT but not necessarily in GPA and LGHSR. Higher users of the study guide tended to be higher on GPA and LGHSR but lower on entrance exam scores. This pattern is more interesting than useful because the correlations were mild at best and allowed little confidence in prediction.

Table 16. Correlations for significant predictors of process variables in the overall group[a]

	LGHPQ		SGUID%[b]		TEXT%[b]	
	r	n	r	n	r	n
GPA	.01	260	.23*	261	-.01	259
LGHSR	-.02	243	-.19*	245	.04	245
ACT	-.17*	213	.09	214	.01	212
MSAT	-.14*	217	.12	218	-.01	216
CONFID	-.01	256	.06	257	.23*	255
PLECT	-.17*	255	-.01	256	-.10	254
PALTT	.05	255	-.10	256	.16*	254
LGTOTH	-.12	258	-.14	259	-.02	257

[a]Sexes and sections combined.

[b]Distributions skewed, correlations are minimal estimates.

*$p < .05$, two-tailed test.

CHAPTER VI. ATTRIBUTE BY TREATMENT INTERACTION RESULTS

The third and final purpose of this investigation is to evaluate PAS using an attribute by treatment interaction model (ATI). In the preceding chapters, the variables were classified as entry, process (study), and outcome. These are renamed in order to be consistent with the ATI literature. All entry variables are referred to as attributes, and all outcomes are referred to as dependent variables. The process variables serve as both depending on the analysis. For example, when the interaction between study hours and treatment is tested in the prediction of SCORE, study hours is an attribute. When an attribute by treatment interaction is tested in the prediction of study hours it serves as a dependent variable.

Several dependent measures were included in the analyses, but academic achievement, as measured by SCORE, received the most attention because it was the most important goal of the zoology course. The attributes were investigated in the order of their review in the literature. Attitudinal outcomes (F1 and F2) and self-reported hours of study time (LGHRPQ) were also studied as dependent variables.

Methods

To statistically test all possible attribute dependent variables pairs would have required over 70 tests and excessive computer time. Instead, the correlational approach described in Chapter II was used as a screening procedure to identify potentially significant interactions for further analysis. The first step was to calculate correlations between all variables for each of four subgroups--PAS males, PAS females (Table 17), TRAD males, and TRAD females (Table 18).[1]

The next step was to informally compare correlations between the attributes and the dependent variable of interest in the four subgroups. For example, with GPA as the attribute and SCORE as the dependent variable, the correlation between these two were compared across treatments. PAS males were compared with TRAD males and PAS females with TRAD females. Those attributes showing relatively large differences in correlations were formally tested with the ATI regression approach. The analysis tested for significant slope differences by examining the contribution of the interaction term after the treatment

[1]The comments on correlational methods in the preceding chapter are relevant here.

Table 17. Pearson correlation matrices for PAS section (males above diagonal, females below diagonal)[a]

	n max	LGTOTH	MAJOR	LGHSR	GPA	MSAT	ACT	HSSCI	CONFID	PLECT	PALTT	HS	FF
n max		36	36	32	34	27	27	32	33	32	33	33	33
LGTOTH	108		-21	-55*	54*	48*	61*	-17	-42*	09	06	05	-32
MAJOR	108	-08		13	-05	-07	-09	10	16	-36*	-13	00	26
LGHSR	95	-16	-11		-70*	-62*	-71*	-24	-02	-30	-12	-14	40*
GPA	105	28*	-20*	-62*		59*	62*	04	-10	16	-18	09	-17
MSAT	82	13	-01	-50*	53*		72*	24	00	10	29	03	-56*
ACT	80	20	-01	-52*	54*	75*		25	14	22	09	19	-50*
HSSCI	97	-11	13	-17	08	21	30*		19	20	07	-32	-19
CONFID	106	-11	36*	-21*	09	27*	34*	43*		14	-04	34	-33
PLECT	106	07	05	07	10	13	04	08	-11		-24	-02	-02
PALTT	105	10	-07	-02	00	-22*	02	-01	-08	-30*		-16	-09
HS	105	06	04	12	-12	09	09	03	20*	-21*	12		-13
FF	106	-19*	03	03	-12	-19	-14	-13	-17	10	-12	-25*	
RAM	105	18	-01	05	01	18	14	11	22*	-19*	14	75*	-83*
TAQ	106	-07	-07	16	-33*	-47*	-42*	-10	-27*	-24*	16	-11	39*
IE	104	00	-01	-02	-15	-33*	-29*	-05	-28*	-18	02	03	34*
LGIELK	104	01	09	14	-19	-35*	-31*	02	-17	-15	09	01	25*
LGIEAC	106	-12	-07	11	-30*	-28*	-29*	-14	-25*	-14	-04	02	10
LGHRPQ	102	-05	-16	10	-05	-29*	-33*	-24*	-11	-22*	07	04	-01
TEXT%[b]	100	05	-20*	18	-04	-10	-11	-24*	09	-17	15	16	-07
SGUID%[b]	102	-09	-28*	-08	11	01	-04	-12	-04	-17	05	01	19
F1	101	11	21*	-08	08	00	11	12	40*	-05	04	11	-30*
F2	101	-15	-11	-02	-09	-08	-15	-15	-27*	07	-04	-14	16
INSTR	101	05	11	10	-03	-01	07	00	07	10	01	-07	-09
SCORE	108	11	13	-51*	63*	42*	43*	08	33*	-07	-15	-09	-04

[a]n's are listed in Appendix B. n max is the maximum n for a variable.

[b]Skewed distribution; correlations are minimal estimates.

*p .05; two-tailed test.

RAM	TAQ	IE	LGIELK	LGIEAC	LGHRPQ	TEXT%[b]	SGUID%[b]	F1	F2	INSTR	SCORE
33	33	33	33	33	33	33	33	31	32	32	36
27	-42*	15	19	-02	-12	-21	-03	-26	20	-14	32
-19	13	-05	-11	09	28	03	16	37*	-42*	13	10
-35	45*	-12	-10	00	-06	02	-23	-01	-09	-19	-44*
18	-50*	-13	-13	-23	13	06	44*	22	-14	12	62*
40*	-38	18	16	09	-21	-30	22	-13	19	-23	54*
46*	-41*	04	04	-03	-20	-15	24	-05	-11	-06	55*
-08	30	22	13	-01	-28	-09	13	14	02	-05	14
44*	14	-13	-11	-02	15	55*	33	48*	-17	03	15
00	20	35*	24	25	-21	08	04	-07	25	-17	20
-03	04	03	09	07	-27	-23	-21	-36	12	-13	-04
68*	-41*	-23	00	-08	-08	14	11	05	04	-12	-09
-81*	38*	01	-04	16	09	-18	-12	-18	-12	-10	-28
	-52*	-14	02	-17	-12	22	16	17	12	01	15
-32*		22	12	13	-01	06	-06	02	08	-09	-20
-20*	49*		83*	75*	-21	-13	-25	-35	37*	-16	19
-16	28*	70*		54*	-17	-16	-22	-27	27	-14	15
-05	39*	64*	37*		-17	-16	-30	-34	40*	-19	-04
02	14	06	16	14		36*	58*	63*	-38*	51*	13
13	01	-17	-02	-18*	45*		31	41*	-16	27	32
-11	12	11	06	13	42*	27*		56*	-23	25	43*
26*	-18	-14	-05	-11	12	10	04		-47*	56*	28
-19	20*	11	20*	17	32*	25*	14	-34*		-43*	-12
02	-07	-19*	-10	-08	-07	-07	-17	42*	-40*		07
-02	-30*	-11	-07	-30	06	-01	23*	29*	-19	04	

Table 18. Pearson correlation matrices for TRAD section (males above diagonal, females below diagonal)[a]

	n max	LGTOTH	MAJOR	LGHSR	GPA	MSAT	ACT	HSSCI	CONFID	PLECT	PALTT	HS	FF
n max		47	47	42	47	40	36	42	35	35	35	34	35
LGTOTH	109		06	10	40*	14	-01	-19	04	-11	05	15	01
MAJOR	109	-19*		-12	07	23	21	27	39*	-12	14	-09	-01
LGHSR	103	00	-17		-47*	-49*	55*	-43*	04	30	20	-13	24
GPA	106	-06	-02	-67*		33*	50*	15	-09	-13	-25	14	-34*
MSAT	91	21*	18	-55*	42*		81*	33*	34	-48*	12	16	-13
ACT	90	03	32*	-67*	57*	72*		31	09	-50*	-12	17	-23
HSSCI	103	-24*	28*	-20*	17	20*	37*		-02	-21	-13	08	-20
CONFID	104	-24*	41*	-14	09	18	19	50*		04	03	-17	01
PLECT	105	10	-14	04	01	09	11	-01	-08		-27	08	21
PALTT	105	18	-16	00	-09	-14	-10	-04	-13	06		11	36*
HS	105	20*	02	-04	-14	-05	06	04	15	01	11		17
FF	104	-01	06	00	08	12	03	-22*	-17	16	05	-10	
RAM	104	14	-01	-02	-17	-12	02	18	21*	-10	04	76*	-74*
TAQ	105	12	-13	23*	-23*	-24*	-36*	-18	-31*	-01	15	-10	15
IE	100	08	-08	05	-01	08	-05	-13	-21*	21*	-03	-14	15
LGIELK	103	04	-04	17	-03	01	-05	-19	-22*	20*	01	-22	14
LGIEAC	101	17	-09	08	-03	-04	-11	-03	-14	17	-11	00	08
LGHRPQ	99	-14	26*	01	-10	-02	-05	-05	07	-10	-09	00	10
TEXT%[b]	99	-01	20*	-11	-01	21	15	07	20*	-06	-02	05	-06
SGUID%[b]	99	-18	10	-24*	31*	18	26*	05	12	10	-18	02	07
F1	99	-15	35*	-07	-01	16	14	25*	52*	03	-21*	30*	-05
F2	99	18	-30*	17	-16	-11	-26*	-36*	-50*	-03	16	-20*	-05
INSTR	97	14	-02	05	-11	05	03	-05	07	15	03	25*	04
SCORE	109	-23*	25*	-48*	51*	28*	45*	44*	41*	07	-21*	-11	04

[a] n's are listed in Appendix B. n max is the maximum n for a variable.

[b] Skewed distribution; correlations are minimal estimates.

*p .05; two-tailed test.

RAM	TAQ	IE	LGIEILK	LGIEAC	LGHRPQ	TEXT%[b]	SGUID%[b]	F1	F2	INSTR	SCORE
34	34	35	35	35	34	35	35	35	35	33	47
06	-05	26	04	05	-05	-03	-18	-28	14	-18	20
-06	-04	-28	-10	02	07	-10	33*	41*	-47*	00	30*
-31	33	38*	31	16	06	24	-19	-31	49*	-23	-32*
39*	-11	-07	-03	-02	-03	-07	17	18	-23	33	64*
20	-36*	-31	-50*	-06	-23	-15	18	14	-28	-05	07
28	-38*	-26	-38*	-10	-20	-22	08	21	-49*	16	23
25	-32	-07	06	-07	14	-29	06	27	-38*	24	26
-13	-14	-20	-28	17	-17	44*	23	52*	-23	-10	14
-11	35*	36*	49*	16	-07	17	00	27	11	20	-05
-23	23	-06	-11	30	-01	07	23	-21	40*	-21	-24
47*	-03	09	-05	17	22	12	-09	09	-02	-05	26
-79*	68*	23	04	26	26	43*	36*	-19	50*	-21	-12
	-60*	-15	-04	-15	-16	-30	-37*	22	-48*	21	27
-16		26	08	26	19	37*	40*	-07	45*	14	-08
-20*	11		71*	42*	-09	06	-33	-17	27	09	-18
-25*	08	70*		09	-08	-20	-22	-04	06	22	00
-04	25*	59*	38*		15	27	05	17	27	11	-06
-07	-04	05	-04	12		08	13	00	12	-10	22
07	-11	03	-02	08	29*		06	19	17	-05	08
-05	-09	06	06	-05	01	01		24	-04	-09	28
28*	-12	-12	-15	02	23*	-03	32*		-59*	49*	46*
-11	25*	22*	24*	05	06	15	-08	-53*		-27	-35*
15	03	17	-21*	00	08	-10	-07	49*	-32*		23
-10	-26*	02	-02	06	12	04	32*	41*	-41*	02	

variable and attribute have been allowed to predict the dependent variable (see Kerlinger and Pedhazur, 1973, and Cronbach and Snow, 1977). At this stage it was discovered that the power of the statistical tests for the males was very low due to small sample sizes. As discussed later in the chapter, the males were dropped from the ATI analyses.

SCORE as a Dependent Variable

Table 19 presents correlations between the attributes and SCORE for the four subgroups. The data provide evidence about sex differences in correlations as well as SECT differences, and attention is briefly turned to that topic. The CONFID correlations showed the greatest sex difference. It was weakly related to SCORE for the males but was a fairly strong predictor for females. CONFID measured interest in zoology and expectancy for success in the course, and hence, two interpretations of the sex difference are warranted: women were better predictors of their success, and being interested in the subject matter was more important for females than for males as far as course achievement was concerned.

Two other variables showed mild tendencies to differ between sexes. Women's SCORE's were somewhat more likely

Table 19. Correlations between all attributes and SCORE[a]

	PAS female (n=80-108)	TRAD female (n=90-109)	PAS male (n=27-36)	TRAD male (n=34-47)
GPA	63*	51*	62*	64*
LGHSR	-51*	-48*	-44*	-32*
ACT	43*	45*	55*	23
MSAT	42*	28*	54*	07
HSSCI	08	44*	14	26
CONFID	33*	41*	15	14
MAJOR	13	25*	10	30
TAQ	-30*	-26*	-20	-08
LGIEAC	-30*	06	-04	-06
PALTT	-15*	-21*	-04	-24
SGUID%[b]	23*	32*	43*	28
LGHRPQ	06	12	13	22
TEXT%[b]	-01	04	32	08
LGTOTH	11	-23*	32	20
PLECT	-07	07	20	-05
HS	-09	-11	-09	26
FF	-04	04	-28	-12
RAM	-02	-10	15	27
IE	-11	02	19	-18
LGIELK	-07	-02	15	00

[a]Abstracted from Tables 17 and 18; exact n's listed in Appendix B.

[b]Skewed distributions, correlations are minimal estimates.

*$p \leq .05$; two-tailed test.

to be affected by test anxiety (TAQ) than male's. Men who had more college experience (LGTOTH) tended to get higher SCORE's than less experienced men, but this was not the case for females. In fact, for TRAD females, those with fewer total credit hours tended to do better.

The main purpose of the chapter concerns SECT, or treatment differences. It is quite apparent that neither GPA nor LGHSR interacted. Their correlations were steady across the groups. Of the remaining general aptitude measures, ACT and MSAT, MSAT showed the greatest discrepancy between PAS and TRAD especially for males. The possible MSAT interaction for males was tested and found not to be statistically significant ($df = 1, 63$; $F = 2.47$; $p > .10$). The interaction was not even close to the .10 level which was surprising given the large difference in correlations. As noted in Chapter II, SECT differences in the variances of an attribute and dependent variable could result in correlation differences but no slope differences. A check on the variances (Tables 6 and 9, in Chapter 4), however, revealed no large differences.

Further consideration revealed that the statistical power of the MSAT test for males was weak because of small sample sizes. Power is the probability that a test will indicate a significant interaction when there is indeed

a true interaction. With MSAT scores available on only 27 PAS and 40 TRAD males, the power of the test with a significance level of .05 was roughly only .30 to .40 (see Cronbach and Snow, 1977). With such dismal chances of detecting a true interaction, the test should never have been performed.[1] Informal inspection of the correlations for the males can help rule out variables and point toward other variables which might be tested in the future with larger samples, but the present sample sizes do not support formal testing.

For the female group which had sample sizes of around 100, the power of the ATI test is much better. With a significance level of .05 the power is between .80 and .90. Continuing the interpretation of correlations in Table 19 for females only, the correlations for ACT were nearly identical, but there was a difference for MSAT. A test for differences in MSAT slopes, however, was not significant ($df = 1, 169$; $F = 0.32$; $p > .10$). The results from GPA, LGHSR, ACT, and MSAT show that general aptitude did not interact with SECT and that PAS was differentially effective for neither lower nor higher aptitude females.

---

[1]The power was low also because the test was for differences in slopes. The power of other statistical tests in the male groups, for example, the ANOVA's and tests of correlations, was more substantial.

The attribute with the largest SECT difference was HSSCI. The number of science semesters taken in high school was directly related to SCORE for the TRAD females but not for PAS females. The interaction was significant ($df = 1, 193$; $F + 8.84$; $p < .01$). The best fit lines for PAS females (SCORE = 55.84 + .46 [HSSCI]) and TRAD females (SCORE = 41.73 + 3.14 [HSSCI]) are plotted in Figure 4.[1] The lines yield a predicted SCORE for a student given her SECT and HSSCI. For example, a woman who had taken two semesters of science scored, on the average, 56 points if she were in PAS and 48 points if in the TRAD SECT. The Johnson-Neyman technique estimates the regions of significance whereby one can state within which ranges of HSSCI there are significant SECT differences. The regions of significance ($p \leq .05$) are indicated in Figure 2. Women with four or less semesters did significantly better in PAS than TRAD. This group constituted 47 PAS and 60 TRAD women. At levels of five, six, and seven semesters, there were no SECT differences. At eight semesters there was a crossover in effectiveness with TRAD significantly higher than PAS females in SCORE. The number of students at this level was quite small, however, with only six students in each section. Efforts were made to explain

[1]Neither line contained a significant nonlinear component.

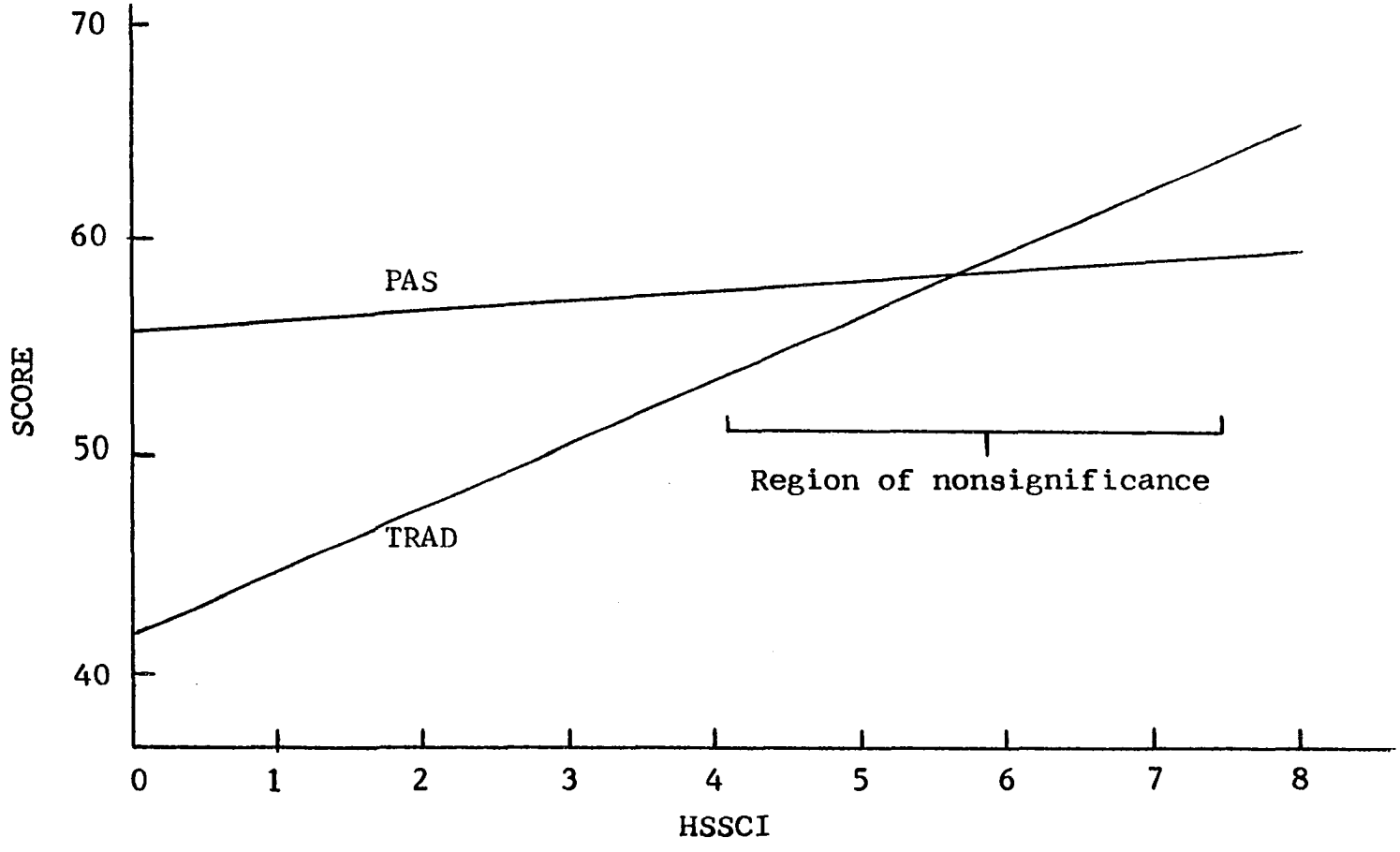


Figure 4. Best fit lines from regression of SCORE on HSSCI for PAS and TRAD female subgroups. Region of nonsignificance calculated from Johnson-Neyman technique

the HSSCI interaction by examining study variables and attitudes. These results are reported later.

The third attribute reviewed in the literature was IE Locus of Control. The full scale IE showed no indications of interaction nor did the luck subscale, LGIELK. The academic subscale, LGIEAC, did show fairly wide SECT differences with the correlations in the expected directions (-.30 for PAS versus .06 for TRAD females). There was a problem with LGIEAC. From preliminary inspection of the correlation matrices in Tables 17 and 18 it was discovered that the correlations between GPA and LGIEAC differed among sections (-.30 for PAS and -.03 for TRAD). This difference may have been due to the brighter PAS students recognizing more the implications of the alternative system in the first part of the quarter and responding more internally. The problem arose because the LGIEAC-SCORE correlation difference may have been an artifact of the LGIEAC-GPA difference. The confoundment was removed in part by controlling with GPA; that is, GPA was allowed to explain SCORE and then the interaction was tested (without GPA controlled: $df = 1, 203$; $F = 5.45$; $p = .02$; with GPA controlled: $df = 1, 197$; $F = 3.05$; $p = .07$). The GPA control reduced the strength of the interaction from a level of statistical significance to

a trend level. There was a tendency for the PAS method to favor the internally academically oriented females while TRAD instruction tended to favor the external females.

For TAQ it was expected that high test anxiety students would perform better in PAS than in TRAD (see Chapter II). This situation would manifest itself in a negative correlation for TRAD students and a zero, or possibly, positive correlation for PAS, but the evidence from Table 19 does not support the expectation. The TAQ-SCORE correlations were negative and of similar magnitude for the females. It is important to consider the measure of course performance in this case. SCORE on the final exam contributed 20% to the final grade and was a "traditional" test in the sense that no retakes were permitted and the student had no options as to when to take it. The correlations show that TAQ was somewhat debilitating to test performance for both groups. Additional information was available for the TAQ-grade correlations which were -.11 for PAS females and -.25 for TRAD females, indicating some, but not large, moderating of the debilitating effects of TAQ in the PAS section.

TAQ was further explored by testing for nonlinearity within sections. The procedure was to simply test the contribution of the square of TAQ as done in the preceding

chapter, but to do it separately for PAS and TRAD female groups. The relationship was linear for PAS but significantly nonlinear for TRAD. This was the case for SCORE and course grade as dependent variables. The best fit curves for SCORE are shown in Figure 5. The equations were SCORE = 67.11 - .13 [TAQ] for PAS and SCORE = 47.38 + .46[TAQ]- .004 $[TAQ]^2$. Visually, the curves are striking. At the extremes of TAQ the PAS exceeded the TRAD females by 10 to 12 exam points, but the sample sizes in the upper and lower regions were quite small. Only 10 and 5 PAS and TRAD females scored below 40, and 5 and 4 above 110 TAQ points. Curvilinear interaction was tested[1] and found not to be statistically significant (df = 2, 199; F = 1.90; $p > .10$).

The correlations for PALTT and the achievement motivation measures (RAM, HS, and FF) were comparable between sections showing no potential for interaction. Hence the interactions were not tested.

Of the remaining attributes in Table 19, only the MAJOR and LGTOTH correlations suggested any possibility of interaction. The MAJOR by SECT interaction did not

---

[1]The contribution of the SECT by TAQ and SECT by TAQ^2 terms was tested after allowing SECT, TAQ, and TAQ^2 main effect to predict SECT.

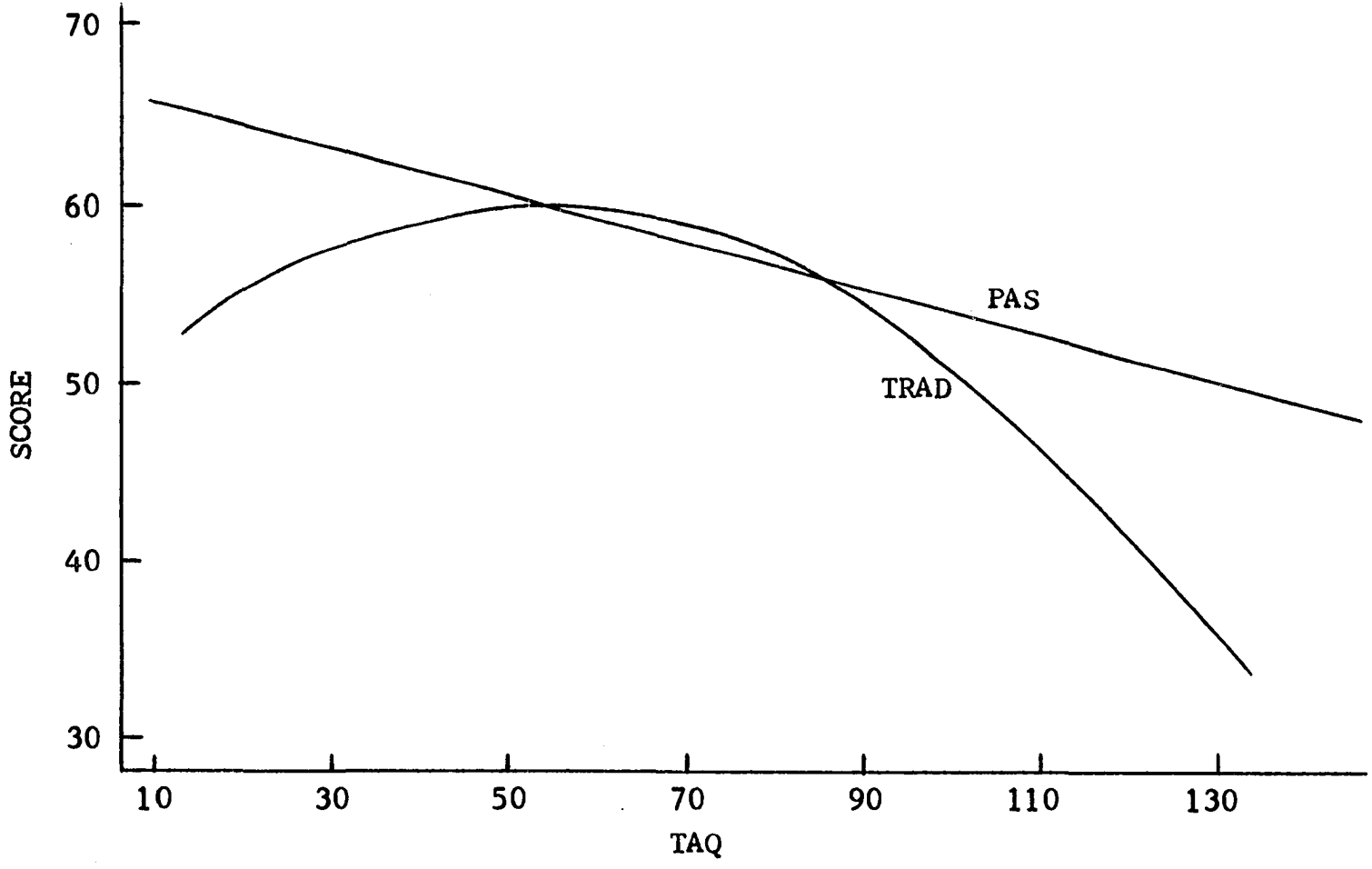


Figure 5. Best fit lines from regression of SCORE on TAQ for PAS and TRAD female subgroups

approach significance (df = 1, 207; F = 1.14; $p > .10$). LGTOTH was confounded with GPA and SECT. For the PAS female section the LGTOTH-GPA correlation was .28 compared to -.06 for TRAD females. When GPA was controlled, the interaction did not approach the .05 level of significance (GPA not controlled: df = 1, 207; F = 6.22; $p < .05$. GPA controlled: df = 1, 206; F = 1.69; $p > .10$).

Student Attitudes as Dependent Variables

Table 20 lists the attributes which correlated significantly with F1 or F2 in either the PAS or TRAD female groups. For the F1 variable (overall satisfaction with course and interest in zoology), the largest correlation difference was for SGUID%[1] and the interaction was significant (df = 1, 196; F = 6.27; $p < .05$). The best fit line for PAS was F1 = 4.23 + .04 [SGUID%], and for TRAD, F1 = 2.12 + .46 [SGUID%]. Although significant the interaction is not interesting: PAS line runs horizontally across the range of SGUID%, and the TRAD line at the lower end of SGUID% starts at lower values of F1 and rises fairly steeply to meet the PAS line at the

[1]The skewed distributions of SGUID% did not provide differences in the correlations because the skewedness was similar in each SECT.

Table 20. Correlations for significant predictors of attitudes[a]

	F1		F2	
	PAS females	TRAD females	PAS females	TRAD females
CONFID	40*	52*	-27*	-50*
MAJOR	21*	35*	-11	-30*
HSSCI	12	25*	-15	-36*
PALTT	04	-21*	-04	16
RAM	26*	28*	-19	-11
TAQ	-18	-12	20*	25*
IE	-14	-12	11	22*
LGIEAC	-05	02	20*	05
LGHPPQ	12	23*	32*	06
SGUID%[b]	04	32*	14	-08
ACT	11	14	-15	-26*

[a]Abstracted from Tables 14 and 15; n's ranged from 97 to 101, exact n's listed in Appendix B.

[b]Skewed distributions, correlations are minimal estimates.

*$p \leq .05$; two-tailed test.

high end of SGUID%. The implications are that females who did not use the study guide were less satisfied with the course in TRAD than in PAS.

A more interesting correlation difference was for PALTT and F1. The interaction was not significant but it did approach significance (df = 1, 191; F = 3.16; p = .07). The best fit lines were F1 = 4.22 + .016 [PALTT] for PAS and F1 = 4.16 - .089 [PALTT] for TRAD females. The lines are graphed in Figure 6. As groups, the PAS and TRAD females were equally satisfied with the course (see ANOVA results in Table 9, Chapter 4), but consideration of student preference for alternative testing (PALTT) revealed that TRAD females tended to be more satisfied than PAS females at lower levels of PALTT with the opposite true at higher levels.

F2 was a measure of dissatisfaction with the testing and grading procedures, and in Chapter IV it was learned that TRAD females were less satisfied than PAS females. Inspection of Table 20 revealed more correlation differences across SECT for F2 than F1, and CONFID, MAJOR, HSSCI, PALTT, LGHRPQ, and SGUID% were formally tested. None was significant at the .05 level, but LGHRPQ was close to significance (df = 1, 196; F = 3.30; p = .07). Figure 7 illustrates the best fit lines for PAS females (F2 = .41 + 1.82 [LGHRPQ]) and TRAD females (F2 = 3.30 +

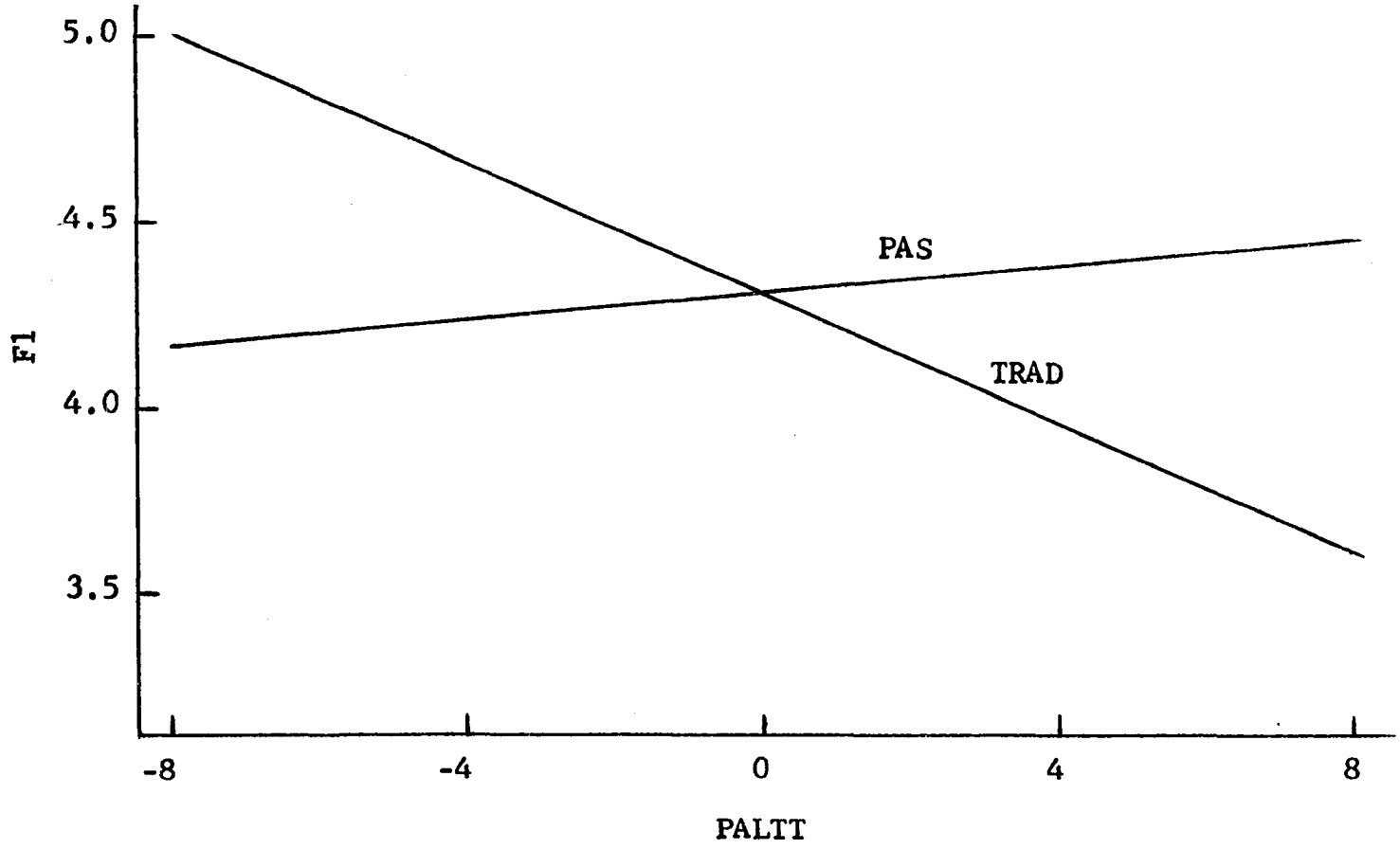


Figure 6. Best fit lines from regression of F1 on PALTT for PAS and TRAD female subgroups

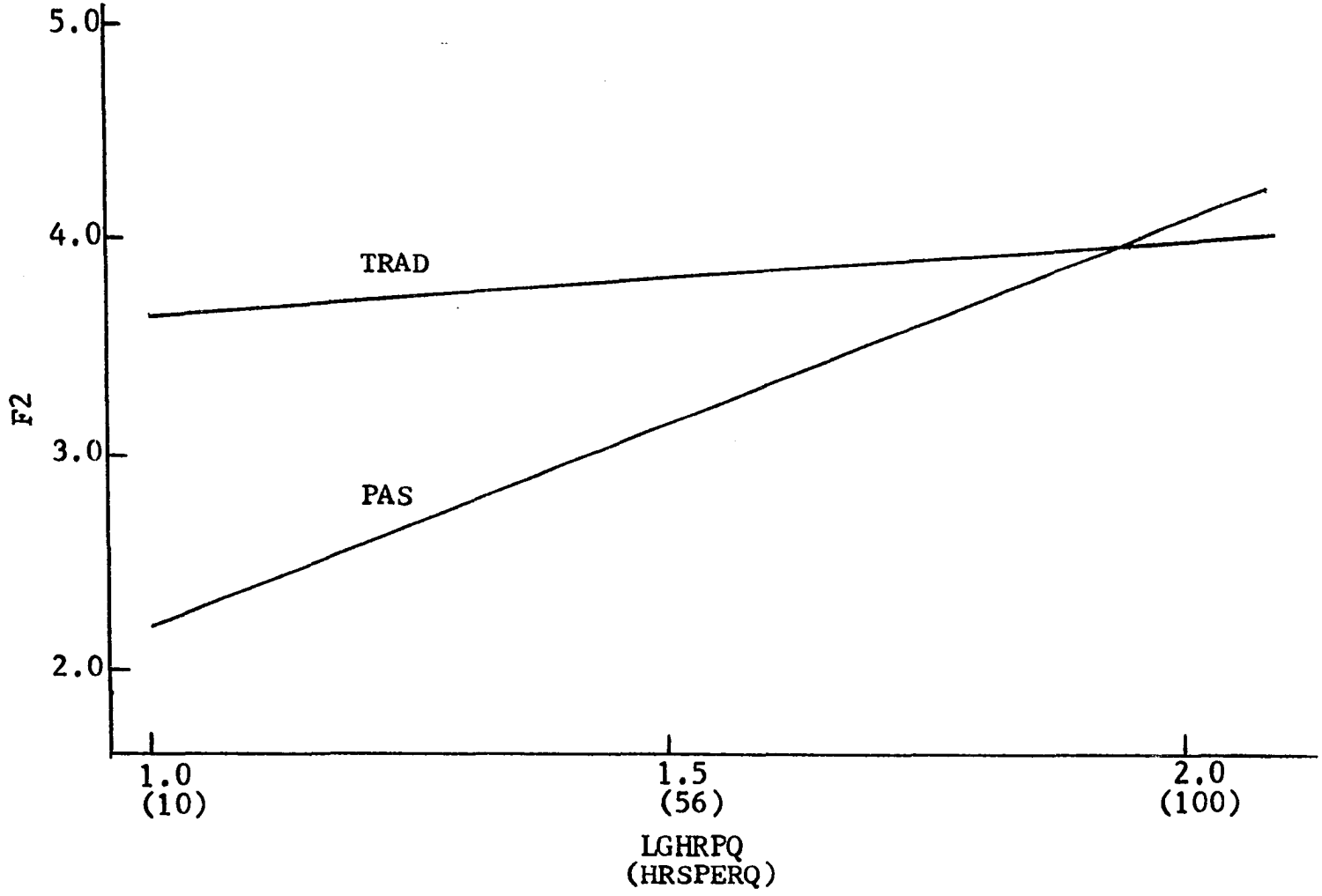


Figure 7. Best fit lines from regression of F2 on LGHRPQ for PAS and TRAD females. HRSPERQ is included for ease of interpretation

.35 [LGHRPQ]). For the most part, TRAD females were more dissatisfied, but at higher levels of LGHRPQ the PAS females were equally dissatisfied.

Study Hours as a Dependent Variable

Table 21 lists the attributes which correlated significantly with HRSPERQ in either the PAS or TRAD female groups. The total reported hours of study time (HRSPERQ) rather than the log transformation (LGHRPQ) was used because the graphs are more meaningful with the actual number of hours.

Table 21. Correlations for significant predictors of HRSPERQ

	PAS female (n = 82 to 102)	TRAD female (n = 90 - 99)
MSAT	-33*	-03
ACT	-41*	-07
MAJOR	-18	21*
HSSCI	-28*	-02
PLECT	-20*	-07

*$p \leq .05$; two-tailed test.

HRSPERQ was more predictable in the PAS than the TRAD group. Of all the attributes, only one, MAJOR, was

significantly related in TRAD. The correlations indicate that women with poorer general and specific background (lower ACT and MSAT scores, nonscience majors, and less high school science) reported more study time than the better prepared women in PAS; in the TRAD section, academic background was essentially unrelated to time spent. Tests for significant interactions confirmed this observation (ACT: $df = 1, 155$; $F = 8.01$; $p < .01$. MSAT: $df = 1, 157$; $F = 4.60$; $p < .05$. MAJOR: $df = 1, 197$; $F = 7.46$; $p < .01$. HSSCI: $df = 1, 182$; $F = 4.22$; $p < .05$). Figure 8 illustrates the ACT and the HSSCI interactions. The PAS females adjusted their study in accordance with their academic background while the TRAD females did not (the lines for them are nearly horizontal). The plots for MSAT and MAJOR (not illustrated) were similar.

The final variable in Table 21, PLECT, showed only a modest difference in correlations, and formal testing revealed no significant difference ($df = 1, 192$; $F = 1.55$; $p > .10$).

Further Investigation of the HSSCI by SECT Interaction

The ATI results on HSSCI revealed that the less prepared females learned more, as measured by SCORE,

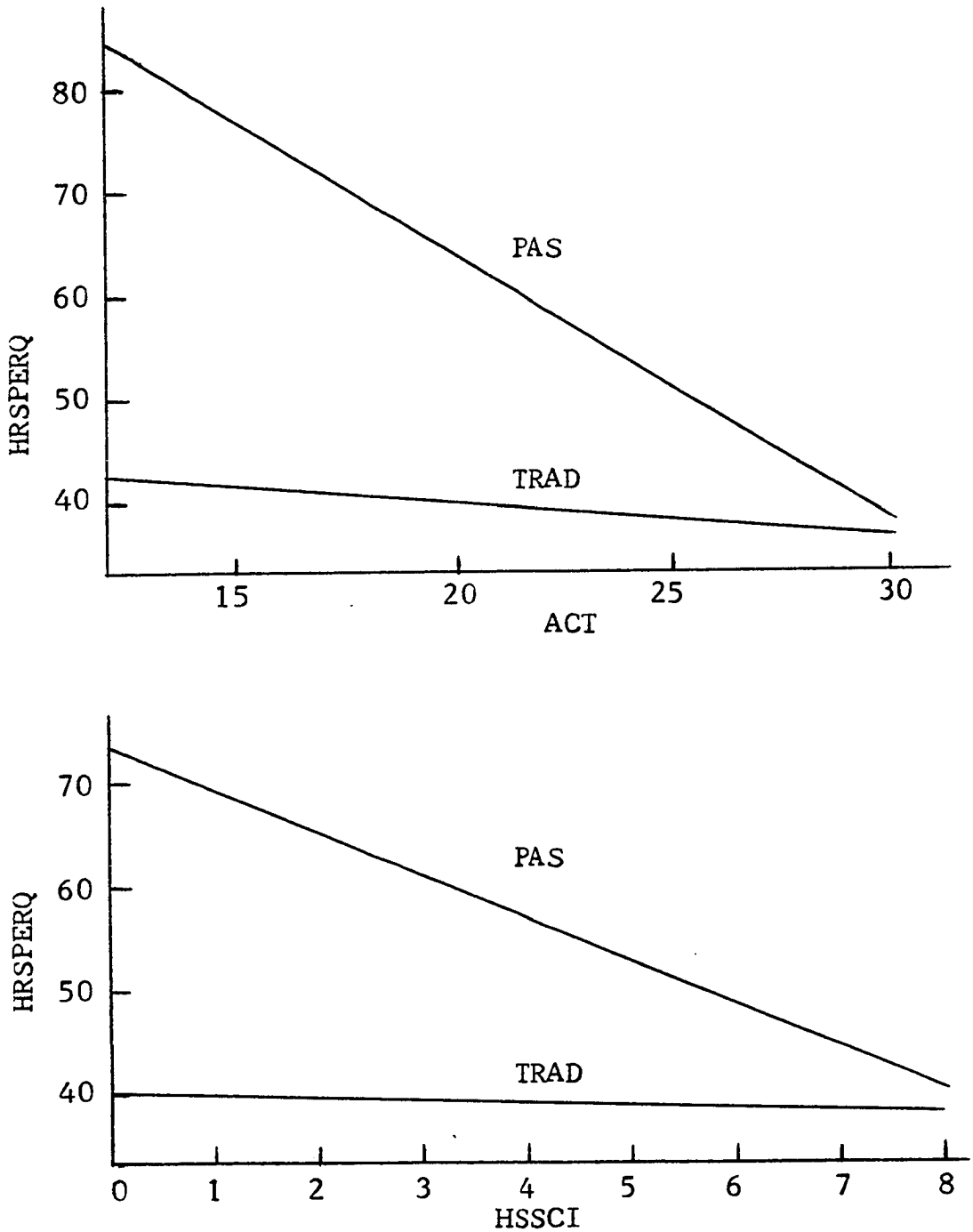


Figure 8. Best fit lines from regressions of ACT and HSSCI on HRSPERQ for PAS and TRAD females

under PAS than TRAD instruction. The interaction is illustrated in Figure 4. Figure 8 shows that they reported spending more time in the PAS section. In an effort to go beyond documenting the interaction and to explain how it happened, the low HSSCI TRAD and PAS females were compared on study variables and attitudes.

The low HSSCI groups were formed by including females who took four or less semesters. This was the region of significant difference on SCORE determined by the Johnson-Neyman technique and it afforded fairly large sample sizes (PAS n = 47, TRAD n = 60). Results from preliminary t-tests indicated no differences on HSSCI or GPA. The PAS did significantly exceed the TRAD females on SCORE (PAS $\bar{x}$ = 57.3, TRAD $\bar{x}$ = 51.9), confirming the Johnson-Neyman finding. Further t-tests demonstrated PAS females reported more study time than the TRAD females, approximately 62 versus 40 hours per quarter. This difference may have been because the PAS females, as an entire group, spent significantly more time than the TRAD females. This possibility was explored by comparing high TRAD and PAS groups. This group consisted of 19 PAS and 14 TRAD female students who took seven or eight semesters of high school science. These two groups reported spending only 36 and 39 hours per quarter, which

is similar to the low HSSCI TRAD females, and more than 20 hours below the average low HSSCI PAS student.

T-tests for SGUID% and TEXT% for the PAS and TRAD low HSSCI groups suggested that the additional time was devoted to the text assignments. The groups were comparable on study guide (PAS x = 72% completion[1], TRAD x = 76%) but differed significantly on TEXT% (PAS $\bar{x}$ = 76%, TRAD $\bar{x}$ = 48%). Again, the text difference may have been due to overall PAS versus TRAD female differences on TEXT%, but the high HSSCI showed more comparable levels of completion of the text assignments (PAS $\bar{x}$ = 58%, TRAD $\bar{x}$ = 46%). The PAS low HSSCI group also showed more agreement that the text reading was necessary to course success (Item 8, Post-Questionnaire, see Appendix B) than did the TRAD low HSSCI group.

Student perceptions of the lectures were also compared, and it was found that the low HSSCI females for both methods held similar views on the necessity of attending lectures (Item 23) and the usefulness of the lectures (Item 25). Two other items about study habits on the post-questionnaire were tested. For Item 13 (cramming was the most effective means of obtaining a high grade), both

---

[1]Raw scores for SGUID% and TEXT% were translated to completion percentages for ease of interpretation.

groups' averages fell at the midpoint between strongly agree and strongly disagree options. It was expected that the low HSSCI PAS females would have reported more adjustment of study habits on the basis of test scores (Item 21), but their self-reports were similar to the TRAD group.

In the area of attitudes, the two low HSSCI groups were not significantly different in their ratings of the overall course (F1) or the instructor (INSTR). The TRAD low HSSCI females were significantly more negative about the testing and grading policies in their course as measured by F2 (TRAD $\bar{x}$ = 4.25, PAS $\bar{x}$ = 3.64). The high HSSCI groups were about equal on this attitude (TRAD $\bar{x}$ = 2.32, PAS $\bar{x}$ = 2.19). On post-questionnaire items, the low HSSCI females showed equal preference for self-paced testing (Item 33), but the PAS group significantly preferred mastery versus normative grading (Item 34). PAS females rated the tests as better measures of knowledge and felt they had more control over their grades than their TRAD counterparts (Items 15, 20).

CHAPTER VII. DISCUSSION AND SUMMARY OF RESULTS

Comparative Evaluation Results

The main research question of Chapter IV was, Which instructional method was more effective in terms of student achievement and attitudes? Before the outcomes were compared, preliminary analyses were done on entry variables. The results indicate that PAS and TRAD students were comparable as they entered the course of instruction and there was no differential dropout. Some interesting effects of PAS on the students occurred early in the quarter. Initial analyses suggested that the explanation of the PAS testing and grading policy to the PAS students caused them to be more positive toward alternative testing procedures and to adopt a more internal perspective of their academic successes and failures. For PAS males there was also an apparent reduction in test anxiety. PAS instruction, however, did not appear to alter expectations for success (CONFID) or motivation (RAM). Both males and females in PAS maintained their positive attitude toward testing throughout the term. For academic locus of control and test anxiety

it is not known how lasting the PAS effects were. The results indicate that the students perceived some fundamental differences between PAS and previous traditional courses.

In comparing student outcomes, the nature of the measure must be kept in mind. The measure of course achievement was performance on a common, final exam. While the exam reliably served its purpose, it was heavily weighted with immediate recall and recognition of vocabulary and facts versus application and synthesis of knowledge. With regard to course achievement neither method was better. PAS students on the average scored two points higher on the 80-item exam (2.5% higher), but the difference was not statistically significant nor can a case for educational significance be made.

This finding is not consistent with the comparative studies of PSI-type approaches to self-paced, mastery instruction, nearly all of which significantly favored PSI over TRAD with an average exam difference of about 9 to 13% (Robin, 1976, and Kulik and Jaksa, 1977).

It would be premature to conclude that PSI is more effective in terms of exam performance than TRAD while PAS is not because PAS has undergone only this one evaluation. The past and present findings do, however,

warrant a brief discussion of the differences between PAS and PSI. The mastery level for a PSI unit is 90 to 100% with grades based on the number of units mastered. In PAS all units must be passed at the 56% level with grades based on the average percentage level across all units. In PSI the results of a unit quiz are immediately discussed with a proctor and the student is either encouraged to make up deficiencies and directed to specific resources, or he is tutored on the unmastered material. In PAS there is no personal feedback, encouragement or tutoring. These two factors may well account for the findings. A third difference lies more in the overall intention of implementation than in the procedures. The PSI approaches were offered as replacements to traditional lecture/discussion, while PAS was designed to supplement traditional instruction with a self-paced, mastery testing and grading procedure. The "out with the old, in with the new" intention of PSI users (both instructors and students) versus the "keep the old, add the new" intention of PAS users may also account for the discrepancy between the present and past comparative findings.

In addition to course achievement, PAS and TRAD students were compared on three attitudes: overall

satisfaction with course and interest in subject matter (F1); general rating of the instructor (INSTR); and dissatisfaction with testing and grading procedures (F2). The sections were comparable on the first two attitudes with only a 2 to 3% difference. On the third attitude, PAS students as a group were significantly less dissatisfied with the testing and grading procedures in their section. The average PAS score on this attitude was 41% of the maximum negative score, and the TRAD average was 48%, yielding a 7% difference. Also, the average PAS students were more positive toward self-paced testing and mastery grading than the TRAD students.

In the area of study patterns, PAS students as a group reported spending significantly more time on the course than TRAD students. The difference was 13 hours for the 10-week term, or 11%. The large section difference was due mostly to the females; they differed by 13% while PAS and TRAD males differed by only 4%. The sections did not differ in reported completion of the study guide (5% difference), but the average PAS student reported reading significantly more of the text assignments (20% difference). As might be expected, they reported more use of test results in adjusting their study behavior.

In spite of the fact that PAS females reported more study time, they did not score higher as a group on the final exam. This inconsistency may be due to the self-reported nature of the variable. Interestingly, while PAS females reported more time, they were less, not more, negative on the testing and grading attitude, part of which measured perceived difficulty (for example, two items were, "This has been a very difficult course" and "I spent too much time on this course for the credit assigned"). This latter item was analyzed separately and there were no significant differences between sections or genders.

In summary, neither PAS or TRAD was more effective in terms of learning the course content, overall student satisfaction, reported interest in subject matter, or rating of the instructor. PAS students were more positive about their testing and grading policy and perceived greater fairness in the procedures. PAS females as a group reported spending more time on the course and completing more of the reading assignments.

A follow-up study of retention would greatly enhance the evaluation. Not only would a comparison of long term learning be informative but also comparisons of changes in attitudes and self-reports of study behavior

and comparisons of performance in subsequent biology courses. In the present study, the students were experienced with traditional methods but naive concerning self-paced, mastery instruction. Further study of the outcomes and study patterns of students who had had previous experience with PAS would also contribute to the findings of the present comparative evaluation.

The major thesis of this paper is that the evaluation methodology should provide for individual differences, particularly when the instructional methodology is based on individual differences. The findings of Chapter IV are grounded in group averages, individual differences were reduced to section (PAS or TRAD) and gender membership. The correlational methodology of Chapter V considers quite minute differences between students, but no consideration is made of the student's section or gender. Chapter VI employs the ATI evaluation design which essentially joins the methodology in Chapter IV with that in Chapter V. Consideration is given to all individual differences including the student's section which results in finer tuning of the findings in Chapters IV and V. The ATI methodology transduces the comparative question from, Which instructional approach is better? to, Which is better for which students?, and it transduces the

correlational question from, Which students do best? to, Which students do best under which method?

Results from the Study of Relationships

Correlations in the overall group (sections and genders combined) were presented in Chapter V to study the relationships between entry, study, and outcome variables.

The general aptitude measures, college GPA, high school rank, ACT, and MSAT were highly related to each other, and in turn, were highly related to course achievement. The single best predictor of final exam SCORE was GPA. It explained 34% of the variability in SCORE confirming the well-documented fact that those who have done well in college courses continue to do well. High school rank was the second highest correlate--those who did well in high school, do well in college.

CONFID, a measure of interest in subject matter and expectancy for success, was another strong correlate of achievement. With GPA it formed the most predictive pair of all possible pairs of variables. Together they explained about 43% of the student differences in SCORE.

With one exception the personality and preference

measures held no practical value in predicting achievement in the overall group. The exception was test anxiety, which was mildly negatively related, and explained, by itself, 6% of the SCORE variance. The factoring of the IE Locus of Control scale did not prove useful in the overall group. The lower correlation between SCORE and preference for alternative testing was expected because half of the students used alternative testing and half did not, but the near zero correlation for preference for lecture was not expected. The utility of the preference variables will be discussed in light of later findings.

Reported hours of study was essentially unrelated to SCORE. By itself it explained only 1% of the SCORE variance. In depth investigation of study hours indicated a somewhat higher relationship for upperclassmen. Other results on study variables indicated that reported effort invested in the study guide tended to be associated with higher grades, but effort invested in the text readings was not.

From the standpoint of predictive utility, the pre- and post-questionnaire variables contributed little to the prediction of SCORE above and beyond the more readily available archival variables. The only variable that retained predictive power was CONFID which explained 5%

of SCORE beyond that of archival measures.

The attitudinal outcomes were less predictable than SCORE. None of the 20 entry and study variables were related to rating of instructor (INSTR). The F1 attitude (overall rating of the course and interest in zoology) and the F2 attitude (dissatisfaction with the testing and grading policy) fared somewhat better. Although SCORE, F1, and F2 were fairly highly interrelated, the predictors of SCORE were not as a rule the predictors of attitudes. For example, GPA, high school rank, and MSAT were not related to either F1 or F2, while the personality scales were mildly predictive. CONFID was the single best predictor for both F1 and F2 explaining 21% and 13% of the student differences in those attitudes. There was a mild trend for harder workers (as measured by LGHRPQ and SGUID%) to hold more positive views toward the course and to be more interested in the subject matter (F1).

In comparison with the cognitive and attitudinal outcomes, the study patterns were essentially unpredictable. Of the 17 entry variables representing a fairly broad spectrum of characteristics, none by itself explained more than 5% of the variance in any of the study variables. This finding together with the comparative findings in Chapter IV indicate that the student's sex and section

(PAS, TRAD) has more influence on study patterns than a host of measures on student characteristics.

In summary, the students who achieved highest in the zoology course were high in general aptitude, were interested in zoology, and expected to do well. For the most part, personality characteristics, preference for instruction, and self-reported study patterns had little to no influence on course achievement in the overall group. No solid conclusions can be drawn, however, until the student's section is considered by examining the relationships within each instructional method. This was the purpose of Chapter VI.

Attribute by Treatment Interaction Evaluation Results

The main purpose of Chapter VI was to evaluate PAS with the ATI model. The research question can be phrased in several ways: Did the PAS and TRAD methods differentially benefit certain types of students in terms of higher achievement? Which types of students achieved the highest in which section? Were the relationships between the student attributes and course achievement the same under the two instructional methods? Which attributes interacted with instructional method?

For the female students, general academic aptitude (GPA, HSR, ACT, and MSAT) clearly did not interact with instructional method in the prediction of course achievement.[1] Neither PAS nor TRAD held special benefit for lower or higher ability students. In both sections the brighter students did considerably better than the less bright students.

The student's level of content specific background, as measured by the number of semesters of high school science, did significantly interact with instructional method. PAS was differentially beneficial for the females with poorer backgrounds. Conversely, TRAD was differentially beneficial for the well-prepared females, but the number of females who benefitted significantly in TRAD was quite small ($n = 6$). Supplemental analyses indicated that PAS fostered more study time in the less prepared PAS females, according to self-reports, and they completed more of the text readings in comparison with the less prepared TRAD females. The evidence suggests that

---

[1]With the exception of this footnote, the results are discussed and summarized for females only. Formal testing of ATI's for males was not done because of the limited statistical power of the tests due to small sample sizes. Informal comparison of the correlations in the PAS and TRAD male groups for course achievement suggested that general aptitude, as measured by ACT and MSAT, might interact with instructional method such that brighter males achieve higher in PAS than in TRAD. This possible interaction requires further investigation with larger sample sizes.

PAS compensated for poor background by encouraging them to put more effort into the course. Although no empirical evidence is available, it would seem that the less prepared females worked harder in PAS than in TRAD because the more frequent testing and feedback in PAS helped them to identify and to remediate their weaker areas. These hypotheses should be tested in future work.

Generalizations, particularly from ATI results, are difficult to make. With this particular implementation of PAS and this particular group of introductory zoology students, less prepared females (less than four semesters of high school science) achieved significantly higher in PAS than TRAD; for average to above average females (five to seven semesters), the sections did not differ while the very well-prepared females (eight semesters) did significantly better in TRAD. These findings do not, however, lead to direct recommendations for matching students with PAS and TRAD. The evidence, particularly if replicated, does provide valuable information to administrators and instructors who must in turn blend that information with a variety of other important factors, such as the costs of instructional methods, the number and nature of potential students, and the availability of human and material resources.

The results for both general aptitude and specific background are consistent with the ATI literature. Self-paced, mastery instruction compensates for specific weaknesses in a student's background but not for lower general ability. The reason for the differences between these two attributes most assuredly lies in the pervasiveness of the deficit. In PAS and PSI-type instruction the weaker student is more able to overcome specific deficits than in traditional instruction, but in neither method is the weaker student able to overcome general deficits in ability to learn.

In an essay on ATI research, Gehlbach (1979) suggested that ideal instructional procedures are those which result in high levels of achievement for all students regardless of individual differences (i.e. methods which are "robust" or "invulnerable" to individual differences). Empirical findings to date reveal that self-paced, mastery instruction has only partially reached this goal. It is robust with regard to differences in content related background but not robust with respect to general aptitude. This goal seems as elusive as it is appealing. After reviewing a large number of laboratory and field studies, Cronbach and Snow (1977 p. 500) drew the following conclusion:

> We once hoped that instructional methods might be found whose outcomes correlate very little with general ability. This does not appear to be a viable hope. Outcomes from extended instruction almost always correlate with pretested ability, unless a ceiling is artificially imposed.

There are several reasons for the discrepancy between the ideal and the present evidence. Rate of learning has been ignored in ATI research on self-paced, mastery instruction. Developers have attempted to provide for differing rates by allowing early and late demonstration of mastery, but the research implementations have restricted time lines because the outcome measures have largely been performance on secure, final exams. Conceivably, if students were allowed several terms in which to master the material, nearly all students would achieve at high levels and general aptitude would not be related to course achievement. It is also possible that prior student experience with self-paced, mastery procedures might moderate the strong influence of ability on course achievement.

Another reason for the discrepancy between Gehlbach's ideal and present state of the art is that the ideal is not obtainable in the real world of higher education. Future research and development on an iterative basis will answer this question. For the present, self-paced, mastery instruction presents one standard against which

to base future instructional robustness to individual differences in content related background.

The full scale of Rotter's Internal-External Locus of Control scale did not interact with instructional treatment. It was not related to course achievement in either of the PAS or TRAD female groups. There was evidence that the academic subscale, identified through factor analysis, did interact at a trend level ($p = .07$) such that the more internal females (those who perceived greater personal control of their academic successes and failures) tended to achieve higher in PAS than in TRAD. This trend must be interpreted cautiously because of a confoundment with GPA (see Chapter VI). The finding clearly indicates the feasibility of further ATI research on locus of control, and the fact that the academic subscale differed in relationship from the full scale, lends support to the validity of using more academic specific scales in educational research.

Test anxiety (TAQ) did not interact with method in the prediction of SCORE. Bearing in mind that SCORE was measured by a no-retake, comprehensive exam, the results reveal that TAQ interfered with exam performance equally for PAS and TRAD females. Stated in other words, PAS did not reduce the detrimental effects of test anxiety

when it came to taking the final exam. There was some moderating effect of PAS on course grade (for PAS, 80% of the grade was based on phase achievement tests; for TRAD, 80% was based on two midterm exams), but the TAQ interaction for grade was not significant.

Achievement motivation (HS, FF, and RAM) did not interact with instructional method. None of the scales were related to SCORE in the PAS or TRAD female groups. The results may be due to inadequate measurement. The scales address achievement but not academic achievement. Therefore, little can be said about the findings of Pascarella (1977). Further ATI study of achievement motivation should employ more educationally relevant measures.

Preference for alternative testing procedures (PALTT) did not interact in the prediction of SCORE which runs contrary to logical reasoning but supports the majority of the literature. The collective results from Chapters V and VI cast serious doubts on the usefulness of both preference variables (PALTT and preference for lecture) in understanding course achievement. Although both were measured with only two-item scales, the problem is not viewed as one of measurement validity. The findings consistently reveal that student preference,

stated at the beginning of the course, was not important with regard to course performance.

ATI investigation with attitudes (F1 and F2) as dependent variables were generally unenlightening. It was found that PALTT was moderately negatively related to overall course satisfaction and interest (F1) in the TRAD female group and unrelated in the PAS group; the interaction approached statistical significance ($p = .07$). There was a trend, then, for PAS to favor (in terms of the F1 attitude) those preferring alternative testing and for TRAD to favor those preferring more traditional testing. The interaction of PALTT in the prediction of the F2 attitude did not approach significance. Although there was a general lack of interactions for the attitudes, this type of research is important, and the present findings should not discourage further study in this area.

The attribute by treatment interactions in the prediction of reported study hours were more informative. Several measures of academic background interacted significantly with instruction; they were ACT and MSAT exam scores, the student's major (science versus nonscience), and semesters of high school science. The PAS students adjusted their effort in accordance with their background (the poorer females reported more study hours while the

more able females reported less), but in TRAD, effort was not adjusted (background did not correlate with study hours). The result for semesters of high school science provides direct support of the compensatory nature of PAS discussed earlier.

The Relationship Between the Purposes

One important benefit of ATI evaluation is that it refines the results from comparative evaluation and the study of relationships. An example from the present study illustrates this fine tuning and traces a major finding through the three purposes for the female students.

In Chapter IV it was learned that PAS and TRAD females entered their respective sections with about the same number of high school science semesters, and they exited, on the average, with comparable levels of achievement. The results from the study of relationships showed that high school science was moderately related to course achievement ($r = .28$ for all females). The ATI results from Chapter VI revealed that PAS and TRAD students did not exit the course with comparable achievement when consideration was given to individual differences in levels of high school science. The low high school science females

in PAS achieved significantly higher than the low science females in TRAD. The very high high school science females learned more under TRAD than PAS procedures. With respect to the finding on relationship, ATI analysis showed that the high school science correlation for the combined female group did not adequately describe the relationship in the separate sections. In PAS, high school science had little influence on exiting achievement ($r = .08$), but in TRAD the influence was quite high ($r = .44$).

In many science disciplines an interaction is viewed as a bothersome phenomenon which obviates a simple interpretation of main effects. In ATI research, a significant interaction can be valuable both in practical terms such as student placement and in understanding the complexities of academic achievement.

Implications for Further Investigation

Follow-up evaluation using the present students would greatly enhance the present findings. Future work could include attention to all three purposes. For example, it would be informative to compare long term retention and application of course content of PAS and TRAD students through follow-up measures of achievement

in life sciences. These measures might include performance in later biology courses as well as essay and objectively scored examinations. Of particular interest is the PAS female group who reported spending more study time than TRAD females. Did their extra effort lead to superior achievement several years hence?

Long term study of achievement is amendable to the individual differences and ATI approaches. An investigatory mode might be used in which all attributes are examined for possible interaction with method in the prediction of long term outcomes. Or specific hypotheses could be formulated on the basis of the present findings and literature on follow-up studies of self-paced, mastery instruction (see Najmaie, 1979). The high school science interaction may not be significant because many of the students will complete some science courses after the present course which will tend to lower the influence high school science has on long term achievement in the TRAD group.

Analysis attitudes and study behavior as recalled several years later would prove valuable in answering such questions as, Do the PAS females recall spending more study time? Do PAS students continue to be more positive than TRAD students toward their respective testing and

grading policies? Do they still show greater preference for alternative testing? How lasting were the effects of PAS on academic locus of control and test anxiety?

Replication with PAS and other versions of self-paced, mastery instruction would contribute to the present results. The present results are based on a relatively short period of instruction of ten weeks. Evaluation over a semester or sequence of semesters might accentuate or erode the treatment, attribute, and ATI effects. Because of low sample sizes, testing for significant ATI for males was not possible; there was some weak evidence that PAS tended to differentially benefit higher aptitude males.

The measure of content related achievement (semesters of high school science) was an adequate measure because most students had had little exposure to science at the college level before entering the zoology course. The validity of this attribute, however, could be improved in other investigations by using pre-test performance, number of prerequsite courses, and achievement in those courses. With regard to general aptitude, the occurrence of a significant ATI does not look promising. Much work lies ahead if one accepts the challenge of developing a robust instructional method which compensates for lower overall

ability at the college level. Cronbach and Snow (1977) and Gehlbach (1979) offer some excellent first steps toward that goal.

Further research with locus of control and achievement motivation is feasible. Significant correlations and ATI's would be more likely to occur if the measures of these attributes were specific to academic situations. It is recommended that consideration be given to dependent variables other than performance on traditional final exams when exploring ATI's for test anxiety. Study of preference for instruction and testing was not productive when the outcome was achievement, but there was a trend for preference to interact with attributes. The validity of preference variables would no doubt improve if students could base their preferences on actual experience with the instruction in question rather than description.

The use of study variable was particularly informative in understanding the high school science interaction and the compensatory nature of PAS. With improved and extended measures of study behavior, this line of inquiry could be very fruitful. The skewedness of the text and study guide variables observed in the present study can be avoided by piloting and adjusting the response scales.

Also, the use of multiple post-questionnaire items might improve the study variables. Perhaps the best approach, however, would be to avoid relying on student memory at the end of the term and have them keep weekly logs of their study behavior. In addition to percentages of completion of the study guide and text assignments, the number of hours spent on these resources should be considered. It is also recommended that measure of lecture attendance and use of lecture notes be included in future work on study patterns.

BIBLIOGRAPHY

Allen, G. J., Gait, L., & Cherney, R. J. Locus of control, test anxiety, and student performance in a personalized instruction course. Journal of Educational Psychology, 1974, 66, 968-973.

Alpert, R., & Haber, R. Anxiety in academic achievement situation. Journal of Abnormal and Social Psychology, 1960, 61, 207-215.

Astin, A., & Panos, R. The evaluation of educational programs. In R. Thorndike (Ed.), Educational measurement. Washington, D. C.: American Council on Education, 1971.

Austin, S. M., & Gilbert, K. E. Student performance in a Keller-Plan course in introductory electricity and magnetism. American Journal of Physics, 1973, 41, 12-18.

Bar-Tal, D., & Bar-Zohar, Y. The relationship between perception of locus of control and academic achievement. Contemporary Educational Psychology, 1977, 2, 181-199.

Bennett, R. M. Student achievement in Biology 101 at Iowa State University. Unpublished doctoral dissertation, Iowa State University, 1970.

Bloom, B. Learning for mastery. Evaluation Comment, 1968, 1(2), 1-13.

Bloom, B. Human characteristics and school learning. New York: McGraw-Hill, 1976.

Boor, M. Dimensions of internal-external control and academic achievement. Journal of Social Psychology, 1973, 90, 163-164.

Born, D. G., & Davis, M. L. Amount and distribution of study in a personalized instruction course and in a lecture course. Journal of Applied Behavior Analysis, 1974, 7, 365-375.

Born, D. G., Gledhill, S. M., & Davis, M. L. Examination performance in lecture-discussion and personalized instruction courses. Journal of Applied Behavior Analysis, 1972, 5, 33-43.

Calhoun, J. F. The relation of student characteristics to performance in a personalized course. Educational Technology, 1975, 15, 16-18.

Callis, V. C. Interactive effects of achievement anxiety, academic achievement, and instructional mode on performance and course attitudes. (Doctoral dissertation, University of Maine, 1977). Dissertation Abstracts International, 1977, 37(8), 4915A.

Carrier, N., & Jewell, D. Efficiency in measuring the effect of anxiety upon academic performance. Journal of Educational Psychology, 1966, 57, 23-26.

Cohen, J., & Cohen, P. Applied multiple regression/correlational analysis for the behavioral sciences. New York: Halsted Press, 1975.

Collins, B. E. Four components of the Rotter internal-external control scale: belief in a difficult world, a just world, a predictable world, and a politically responsive world. Journal of Personality and Social Psychology, 1974, 29, 381-388.

Cronbach, L. J. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684.

Cronbach, L. J., & Snow, R. E. Aptitudes and instructional methods. New York: Irvington Publishers, 1977.

Cross, M., & Semb, G. A personalized university course in nutrition. Journal of Nutrition Education, 1975, 4, 149-151.

Cross, M., & Semb, G. An analysis of the effects of personalized instruction on students at different initial performance levels in an introductory college nutrition course. Journal of Personalized Instruction. 1976, 1, 47-50.

Daniels, R. L., & Stevens, J. P. The interaction between the internal-external locus of control and two methods of college instruction. American Educational Research Journal, 1976, 13, 103-113.

Davis, M. L. Mastery test proficiency requirement affects mastery test performance. In J. M. Johnston (Ed.), Behavior research and technology in higher education. Springfield, Illinois: Charles C. Thomas, 1975.

Dolphin, W. D., Franke, R. G., Covert, F. G., & Jorgensen, C. D. The phase achievement system: An instructional management system for large enrollment lecture sections. American Institute of Biological Sciences Educational Review, 1973, 3, 24-27, 40-42.

Dorsel, T. The effect of preference for method and type on comprehension of verbal material. Journal of Experimental Education, 1976, 45(1), 30-33.

Dubin, R., & Taveggia, T. The teaching-learning paradox. Eugene, Oregon: University of Oregon Press, 1968.

DuNann, D. H., & Weber, S. J. Short-and-long-term effects of contingency managed instruction on low, medium, and high GPA students. Journal of Applied Behavior Analysis, 1976, 9, 375-376.

Eilersen, H. J. Jr. Locus of control and achievement and participation in structured and unstructured introductory psychology courses. Unpublished doctoral dissertation, University of Washington, 1972.

Eisenman, R., & Platt, J. J. Birth order and sex differences in academic achievement and internal-external control. Journal of General Psychology, 1968, 78, 279-285.

Fernald, P., & DuNann, D. Effects of individualized instruction upon low- and high-achieving students, study behavior, and student's evaluation of mastery. Journal of Experimental Education, 1975, 43, 27-34.

Flanagan, J. Project Talent: the American high school student. Pittsburgh: University of Pittsburgh Press, 1964.

Franklin, T. E. The prediction of achievement and time spent in instruction in a self-paced individualized course. Menomie, Wisconsin: University of Wisconsin--Stout, 1976. (ERIC Document No. ED 124 043)

Gaynor, J., & Millham, J. Individual differences and course performance. British Journal of Educational Psychology, 1975, 45, 286-292.

Gehlbach, R. Individual differences: implications for instructional theory, research, and innovation. Educational Researcher, 1979, 8(4), 8-14.

Gindler, H., Marosz, W., & Romano, A. Who benefits most from individualized instruction in mathematics? Journal of Personalized Instruction, 1977, 2, 204-208.

Gozali, H., Cleary, A. T., Walster, W. G., & Gozali, J. Relationship between the internal-external control construct and achievement. Journal of Educational Psychology, 1973, 64, 9-14.

Hays, W. L. Statistics. New York: Holt, Rinehart, and Winston, 1963.

Henneberry, J. K. Initial progress rates as related to performance in a personalized system of instruction. Teaching of Psychology, 1976, 3, 178-181.

Hess, J. H. Keller Plan instruction: Implementation problems. In J. Sherman (Ed.), Personalized systems of instruction, 41 germinal papers. Menlo Park: W. A. Benjamin Inc., 1974.

Hjelle, I. A. Internal-external control as a determinant of academic achievement. Psychological Reports, 1970, 26, 326.

Hohn, R. L., DesLauriers, M., & Deaton, W. Learners' characteristics and performance effects in self-paced instruction. Psychological Reports, 1977, 40, 1011-1012.

Hopper, H. H. Predictors of college success. Pulaski, Virginia: Virginia Community College, 1968. (ERIC Document No. ED 024 374)

Iowa State University Testing and Evaluation Center, Comments regarding the validity of the placement testing program. Unpublished reports, Iowa State University Testing and Evaluation Center, Carver Hall, Ames, Iowa, 1975, 1976, 1977, 1978.

Jacko, E. J. Lecture versus PSI on individuals with differing achievement anxiety and academic achievement. Unpublished doctoral dissertation, University of Tennessee, 1974.

Johnson, W. G., & Croft, R. G. Locus of control and participation in a personalized system of instruction course. Journal of Educational Psychology, 1975, 67, 416-421.

Johnsten, T. D. A study of achievement in general college biology at the Univeristy of Nebraska as related to high school backgrounds in science and mathematics. (Doctoral dissertation, University of Nebraska, 1967). Dissertation Abstracts, 1967, 28, 1989A.

Kaemmerer, W. F., & Schwebel, A. I. Factors of Rotter internal-external scale. Psychological Reports, 1976, 39, 107-114.

Keller, F. S. Goodbye, teacher . . . . Journal of Applied Behavior Analysis, 1968, 1, 79-89.

Keller, J., Goldman, J., & Sutter, J. Locus of control in relation to academic attitudes and performance in a personalized system of instruction course. Journal of Educational Psychology, 1978, 70, 414-421.

Kerlinger, F., & Pedhazur, E. Multiple regression in behavioral research. New York: Holt, Rinehart, and Winston, Inc., 1973.

Klauw, C. F., & Plomp, T. Individualized study systems in theory and practice. Higher Education, 1974, 3, 213-230.

Kulik, C., & Kulik, J. PSI and the mastery model. In B. Green (Ed.), Personalized instruction in higher education: Proceedings of the second national conference. Washington, D. C.: Center for Personalized Instruction, 1976.

Kulik, J. A., & Jaksa, P. PSI and other educational technologies in college teaching. Educational Technology, 1977, 17(9), 12-19.

Kulik, J. A., Kulik, C., & Carmichael, K. The Keller Plan in science teaching. Science, 1974, 183, 379-383.

Kulik, J. A., Kulik, C., & Milholland, J. E. Evaluation of an individualized course in psychological statistics. In R. S. Ruskins & S. F. Bono (Eds.), Proceedings of the First National Conference. Washington, D. C.: Center for Personalized Instruction, 1974.

Kulik, J. A., Kulik, C. C., & Smith, B. B. Research on the personalized system of instruction. Programmed Learning and Educational Technology, 1976, 13, 23-30.

Latta, R. Factor analysis of an achievement motivation scale. Unpublished report, Department of Psychology, University of New Hampshire, Durham, New Hampshire, 1975.

Latta, R., Dolphin, W., & Grabe, M. Individual differences model applied to instruction and evaluation of large college classes. Journal of Educational Psychology, 1978, 70, 960-970.

Linn, R., & Slinde, J. The determination of the significance of change between pre- and post testing periods. Review of Educational Research, 1977, 47, 121-151.

Lipton, L. A. Relationships among IE control, academic ability, attitude toward science, and achievement in introductory college physics classes. (Doctoral dissertation, Temple University, 1976). Dissertation Abstracts International, 1976, 37(4), 2095A.

Ludwig, L. M. Aptitude and mastery learning in a personalized, taxonomy-based community college social science course. (Doctoral dissertation, Kent State University, 1977). Dissertation Abstracts International, 1977, 37(12), 7512A.

Massari, D. J., & Rosenblum, D. C. Locus of control, interpersonal trust and academic achievement. Psychological Reports, 1972, 31, 355-360.

Mehrabian, A. Measures of achieving tendency. Educational and Psychological Measurement, 1969, 29, 445-451.

Mirels, H. L. Dimensions of internal versus external control. Journal of Consulting and Clinical Psychology, 1970, 34, 226-228.

Morris, C., & Kimbrell, G. Performance and attitude effects of the Keller method in an introductory psychology course. Psychological Record, 1972, 22, 523-530.

Morris, C., & Kimbrell, G. Individual differences and PSI: A reanalysis. Journal of Personalized Instruction, 1977, 2, 47-49.

Najmaie, M. H. Longitudinal effects of the self-paced instructional method in an introductory university biology course. Unpublished doctoral dissertation, Iowa State University, 1979.

Nation, J. R., & Roop, S. S. A comparison of two mastery approaches to teaching introductory psychology. Teaching of Psychology, 1975, 2, 108-111.

Nazzaro, J. R., Todorov, J. C., & Nazzaro, J. W. Student ability and individualized instruction. Journal of College Science Teaching, 1972, 2, 29-30.

Nie, N., Hull, C., Jenkins, J., Steinbrenner, K., & Bent, D. Statistical package for the social sciences. New York: McGraw-Hill, 1975.

Nunnally, J. Psychometric theory. New York: McGraw-Hill, 1967.

Parent, J., Forward, J., Canter, R., & Mohling, J. Interactive effects of teaching strategy and personal locus of control on student performance and satisfaction. Journal of Educational Psychology, 1975, 67, 764-769.

Pascarella, E. Student motivation as a differential predictor of course outcomes in personalized system of instruction and conventional instructional methods. Journal of Educational Research, 1977, 71, 21-26.

Pascarella, E. T. Interactive effects of prior mathematics preparation and level of instructional support in college calculus. American Educational Research Journal, 1978, 15, 275-285.

Prociuk, T. J., & Breen, L. J. Internal-external control, test anxiety, and academic achievement. Additional data. Psychological Report, 1973, 33, 563-566.

Reynolds, C. H., & Gentile, J. R. Performance under traditional and mastery assessment procedures in relation to students' locus of control: A possible aptitude by treatment interaction. Journal of Experimental Education, 1976, 44(4), 55-60.

Robin, A. L. Behavioral instruction in the college classroom. Review of Educational Research, 1976, 46, 313-354.

Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 1966, 80(1, Whole No. 609).

Rotter, J. B. Some problems and misconceptions related to the construct of internal versus external control of reinforcement. Journal of Consulting and Clinical Psychology, 1975, 43, 56-67.

Ryan, B. A. PEI: Keller's personalized system of instruction: An appraisal. Washington, D. C.: American Psychological Association, 1974.

Sarason, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. Journal of Abnormal and Social Psychology, 1958, 56, 339-344.

Schimpfhauser, F., Horrocks, L., Richardson, K., Alben, J., Schumm, D., & Sprecher, H. The personalized program of instruction as an adoptable alternative within the traditional structure of medical basic sciences. In R. S. Ruskins & S. F. Bono (Eds.), Proceedings of the First National Conference. Washington, D. C.: Center for Personalized Instruction, 1974.

Semb, G. An experimental analysis of personalized instruction: A case study from the Land of Oz. In B. A. Green, Jr. (Ed.), Proceedings of the Second National Conference. Washington, D. C.: Center for Personalized Instruction, 1976.

Somon, M. L. Performance and its relationship to self-concept of academic ability, LOC orientation and choice of instructional plan. (Doctoral dissertation, West Virginia University, 1976). Dissertation Abstracts International, 1976, 37(4), 1967A.

Szabo, M., & Feldhusen, J. Success in an independent study science course at the college level as related to intellective, personality, and biographical variables. Journal of Research in Science Teaching, 1971, 8, 225-229.

Tamir, P. An analysis of certain achievements and attitudes of Cornell students enrolled in introductory biology with special reference to their high school preparation. (Doctoral dissertation, Cornell University, 1968). Dissertation Abstracts International, 1969, 29(11), 3924A.

Tortora, G., & Anagnostakos, N. Principals of anatomy and physiology. San Francisco: Canfield Press, 1975.

VanVerth, J. E., & Dinan, F. J. A Keller Plan course in organic chemistry. In R. S. Ruskins & S. F. Bono (Eds.), Proceedings of the First National Conference. Washington, D. C.: Center for Personalized Instruction, 1974.

Wagner, J. Summary of selected results from year one of the PAS CAUSE project. Unpublished report, Biology Department, Iowa State University, Ames, Iowa, 1977.

Warehime, R. G. Generalized expectancy for locus of control and academic performance. Psychological Reports, 1972, 30, 314.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1971.

Wood, W. S., & Wylie, R. G. Individualized systems of instruction are better . . . for whom? In J. M. Johnston (Ed.), Behavior research and technology in higher education. Springfield, Illinois: Charles C. Thomas, 1975.

ACKNOWLEDGMENTS

In addition to the financial support from a National Science Foundation CAUSE grant (#SER 76-16100), many people have helped in this work. The advice of Dr. Trevor Howe, Dr. Richard Warren, Dr. Anton Netusil, and Dr. Leroy Wolins has been much appreciated. I would also like to thank my committee members, Dr. Ross Engel, Dr. Wilbur Layton, Dr. Warren Dolphin, Dr. Dan Reschly, Dr. Rex Thomas and Dr. John Menne, for their assistance in designing and reporting the study. The guidance of my major professor, Dr. J. S. Ahmann, has been invaluable throughout my entire graduate education.

Dr. Dolphin deserves special recognition for the many hours he generously gave to me. I sincerely thank him. Without his ideas and encouragement, this research would not have been possible.

Finally, I thank my wife, Julie, and children, Beth and Dann. They not only told me, but showed me, that the sun also rises.

APPENDIX A.

PRE-QUESTIONNAIRE (ATTITUDE SURVEY) AND POST-QUESTIONNAIRE (CAUSE QUESTIONNAIRE)

ATTITUDE SURVEY

Enter your name and social security number on the IBM sheet where it says NAME and IDENTIFICATION NUMBER. Be sure to indicate sex with M or F. Please use the following scale to indicate the degree of your agreement or disagreement with each of the opinion questions on the following pages. Mark your answers on the IBM answer sheet. Be sure the number of the statement agrees with the number on the answer sheet. Make your marks heavy and black. Erase completely any answer you wish to change. Do not leave any blank spaces. Do not use the zero (0) answer.

9 = very strong agreement
8 = strong agreement
7 = moderate agreement
6 = slight agreement
5 = neither agreement nor disagreement
4 = slight disagreement
3 = moderate disagreement
2 = strong disagreement
1 = very strong disagreement

These questionnaires will be analyzed by an independent agency after course grades are awarded. Please be honest in your response.

1. I prefer independent study and discussion sections instead of lecture classes.

2. I feel that I am responsible for determining my progress and grade in this course.

3. Students should be allowed to take tests when they are prepared even though this might be at a slower or faster pace than that set by the instructor.

4. Since I have a good background in the sciences, I expect I will do well in this course.

5. I see no benefit in taking this course in biology but I must since it is a requirement for graduation.

6. I have always been interested in biology.

7. My friends told me that this course has a reputation for being difficult.

8. A detailed study guide with sample examination questions helps a student organize his study effort for a course.

9. I do not plan to spend a lot of time studying for this course this quarter.

10. I prefer to take courses in large lecture sections.

11. I would prefer to have tests once a week rather than twice a quarter.

12. Grades on tests and the course grade should always be based on "the curve" rather than pre-set standards.

13. If students are allowed to retake tests to change grades, academic standards are lowered.

14. While taking an important examination, I perspire a great deal.

15. I get to feel very panicky when I have to take a surprise exam.

16. I study longer and harder than other students.

17. During tests, I find myself thinking of the consequences of failing.

18. After important tests I am frequently so tense that my stomach gets upset.

19. If I could possibly avoid it, I would never want to take an intelligence test.

20. While taking an important exam I find myself thinking of how much brighter the other students are than I am.

21. I freeze up on things like intelligence tests and final exams.

22. If I were to take an intelligence test I would worry a great deal before taking it.

23. During course examinations, I find myself thinking of things unrelated to the actual course material.

24. During a course examination, I frequently get so nervous that I forget facts I really know.

25. I usually get depressed after taking a test.

26. Even though it serves no purpose, I spend a lot of time thinking of ways to avoid taking tests.

27. I have an uneasy, upset feeling before taking a final examination.

28. When taking a test, my emotional feelings do not interfere with my performance.

29. Getting a good grade on one test doesn't seem to increase my confidence on the second.

30. While taking an important test I have on occasion noticed that my heart is beating very fast.

31. After taking a test I always feel I could have done better than I actually did.

32. I sometimes feel my heart beating very fast during important tests.

33. I would be willing to stake my continuance in school on the outcome of a group intelligence test which is known to be reliable.

34. If I knew I was going to take an intelligence test I would feel confident and relaxed beforehand.

35. (male)* I would rather work on a task where I, alone, am responsible for the final product than one in which many people contribute to the final product.

35. (female)* I more often attempt difficult tasks that I am not sure I can do than easier tasks I believe I can do.

---

*Some of the items on the Resultant Achievement Motivation Scale differed for the genders.

36. (male)* I more often attempt difficult tasks that I am not sure I can do than easier tasks I believe I can do.

36. (female)* I would rather do something at which I feel competent and relaxed than something which is challenging and difficult.

37. (male)* If I am not good at something, I would rather keep struggling to master it than move on to something I may be good at.

37. (female)* If I am not good at something I would rather keep trying to master it rather than move on to something I may be good at.

38. I would prefer a job which is important, difficult, and involves a 50% chance of failure to a job which is womewhat important but not difficult.

39. (male)* I worry more about getting a bad grade than I think about getting a good grade.

39. (female)* I would rather have a job in which my role is clearly defined by others and rewards could be higher than average than a job in which my role is to be defined by me and my rewards are average.

40. (male)* I think that I hate losing more than I love winning.

40. (female)* My strongest feelings are aroused more by fear of failure than by hope of success.

41. For me, the pain of getting turned down after a job interview is greater than the pleasure of getting hired.

42. I am more unhappy about doing something badly than I am happy about doing something well.

---

*Some of the items on the Resultant Achievement Motivation Scale differed for the genders.

For the following questions use these directions. This is a questionnaire to find out the way in which certain important events in our society affect different people. Each item consists of a pair of alternatives numbered 1 or 2. Please select the one statement of each pair (and only one) which you more strongly believe to be the case as far as you're concerned. Be sure to select the one you actually believe to be more true rather than the one you think you should choose or the one you would like to be true. This is a measure of personal belief: obviously there are no right or wrong answers. Do not use zero or numbers 3, 4 etc.

43. (1) Many of the unhappy things in people's lives are partly due to bad luck.
    (2) People's misfortunes result from the mistakes they make.

44. (1) One of the major reasons why we have wars is because people don't take enough interest in politics.
    (2) There will always be wars, no matter how hard people try to prevent them.

45. (1) In the long run people get the respect they deserve in this world.
    (2) Unfortunately, an individual's worth often passes unrecognized no matter how hard he tries.

46. (1) The idea that teachers are unfair to students is nonsense.
    (2) Most students don't realize the extent to which their grades are influenced by accidental happenings.

47. (1) Without the right breaks one cannot be an effective leader.
    (2) Capable people who fail to become leaders have not taken advantage of their opportunities.

48. (1) No matter how hard you try some people just don't like you.
    (2) People who can't get others to like them don't understand how to get along with others.

49. (1) I have found that what is going to happen will happen.

(2) Trusting to fate has never turned out as well for me as making a decision to take a definite course of action.

50. (1) In the case of the well-prepared student there is rarely if ever such a thing as an unfair test.
(2) Many times exam questions tend to be so unrelated to course work that studying is really useless.

51. (1) Becoming a success is a matter of hard work, luck has little or nothing to do with it.
(2) Getting a good job depends mainly on being in the right place at the right time.

52. (1) The average citizen can have an influence in government decisions.
(2) This world is run by the few people in power, and there is not much the little guy can do about it.

53. (1) When I make plans, I am almost certain that I can make them work.
(2) It is not always wise to plan too far ahead because many things turn out to be a matter of good or bad fortune anyhow.

54. (1) In my case getting what I want has little or nothing to do with luck.
(2) Many times we might just as well decide what to do by flipping a coin.

55. (1) Who gets to be the boss often depends on who was lucky enough to be in the right place first.
(2) Getting people to do the right thing depends upon ability, luck has little or nothing to do with it.

56. (1) As far as world affairs are concerned, most of us are the victims of forces we can neither understand nor control.
(2) By taking an active part in political and social affairs the people can control world events.

57. (1) Most people don't realize the extent to which their lives are controlled by accidental happenings.

(2) There really is no such thing as "luck".

58. (1) It is hard to know whether or not a person really likes you.
(2) How many friends you have depends upon how nice a person you are.

59. (1) In the long run the bad things that happen to us are balanced by the good ones.
(2) Most misfortunes are the result of lack of ability, ignorance, laziness, or all three.

60. (1) With enough effort we can wipe out political corruption.
(2) It is difficult for people to have much control over the things politicians do in office.

61. (1) Sometimes I cannot understand how teachers arrive at the grades they give.
(2) There is a direct connection between how hard I study and the grades I get.

62. (1) Many times I feel that I have little influence over the things that happen to me.
(2) It is impossible for me to believe that chance or luck plays an important role in my life.

63. (1) People are lonely because they do not try to be friendly.
(2) There is not much use in trying too hard to please people, if they like you, they like you.

64. (1) What happens to me is my own doing.
(2) Sometimes I feel that I do not have enough control over the direction my life is taking.

65. (1) Most of the time I cannot understand why politicians behave the way they do.
(2) In the long run the people are responsible for bad government on a national as well as on a local level.

66. Indicate what grade you expect to earn in this course: (1) A; (2) B; (3) C; (4) D; (5) F; (6) I.

COURSE QUESTIONNAIRE

Print your name and social security number in the appropriate boxes on the accompanying answer sheet. Blacken the spaces corresponding to the letters and numbers in the columns beneath. Fill in the course and section number.

We feel that the real use of an evaluation is to improve the course for future students. For that reason, we would like to use your experience in this course as the basis for evaluations which may strongly influence how we structure the course in the future. These questionnaires will be analyzed by an independent agency after course grades are awarded. Please be honest in your response.

Questions:

1. This course is _____in my curriculum.
   (1) required; (2) recommended; (3) just an elective

2. How many hours per week did you spend on this course outside of class?
   (1) 1 hr; (2) 2-4 hr; (3) 5-6 hr; (4) 7-8 hr;
   (5) 9-10 hr; (6) more than 10 hr.

3. How many lectures did you miss during the quarter?
   (1) 0 or 1; (2) 2; (3) 3; (4) 4; (5) 5; (6) 6;
   (7) 7; (8) 8; (9) 9 or more.

4. What grade do you think you should receive in this course?
   (1) A; (2) B; (3) C; (4) D; (5) F; (6) I.

5. Approximately what fraction of the suggested textbook readings did you read during the quarter?
   (1) 20% or less; (2) 21-40%; (3) 41-60%;
   (4) 61-80%; (5) 81-100%.

6. Approximately what fraction of the questions in the study guide did you conscientiously answer?
   (1) 20% or less; (2) 21-40%; (3) 41-60%;
   (4) 61-80%; (5) 81-100%.

7. Given the material that was presented in this class, the size of the class was:
(1) too large; (2) about the right size;
(3) could have been larger.

Opinions:

Please use the following continuous scale to indicate the degree of your agreement or disagreement with each of the opinion questions which follow. Mark your answers on the IBM answer sheet. Be sure the number of the statement agrees with the number on the answer sheet. Make your marks heavy and black. Erase completely any answer you wish to change. Do not leave any blank spaces. Do not use response zero (0).

1	2	3	4	5	6	7	8	9
(Strong disagreement = 1)				(Neutral = 5)				(Strongly agree = 9)

8. I felt that I had to do all of the assigned readings in order to do well in this course.

9. Compared to other courses at ISU, the tests in this course were more threatening.

10. Too much emphasis was placed on testing and grades in this course.

11. During the course, my interest in biology developed to the point that I wanted to spend more time than I had originally expected.

12. This course contains a lot of busy work that is not related to understanding concepts and principles.

13. In this course, cramming for tests was the most effective means of obtaining a high grade.

14. I am generally pleased with the text book required in this course.

15. The tests were an adequate measure of my knowledge and will allow the instructor to assign me the grade I deserve.

16. The grade standards in this course are too high.

17. I felt that I had to answer all of the study guide questions in order to do well in this course.

18. I think this is one of the better courses I have had in science.

19. If video tapes supplemental to the lectures were available in the library, I would use them as an information source.

20. I felt that I could determine my grade in this course more than in most courses at ISU.

21. I adjusted my study habits during the course according to the scores I received on tests.

22. I perceived that I had freedom in this course to arrange my study schedule to accommodate my interests and the demands placed on me by other courses.

23. Frequent attendance in this class is essential to good learning.

24. Compared to other courses I took this quarter, I spent too much time on this course for the credit assigned.

25. I felt the lectures were useful.

26. This course forced me to regard myself as being unable to comprehend the basic concepts of biology.

27. I am satisfied with the overall organization, administration, and instruction offered by this course.

28. I felt the study guide was helpful.

29. Test results were useful to me in planning my studying for this course.

30. The number of exams was not adequate to test my understanding and keep my interest.

31. This has been a very difficult course.

32. My level of interest in the biological sciences has increased as a result of taking this course.

33. I would prefer to take tests at my own pace rather than as instructor-scheduled, required midterms.

34. I feel that course grades should be based on "the curve" rather than based on pre-set standards.

35. The course allowed me to pursue in depth understanding in areas that personally interested me.

36. This course has stimulated my desire to take additional biology courses.

37. I feel that I have mastered the relevant content of this courses.

38. My final grade will be limited because I lack a science background.

39. I would recommend that other students take this course.

Opinion of Instructor:

40. The instructor did not interpret abstract ideas and theories clearly.

41. The instructor contributed to my interest in his subject.

42. The instructor has helped broaden my interests.

43. The instructor has increased my skills in thinking.

44. The instructor does not stress important material.

45. The instructor makes good use of examples and illustrations.

46. The instructor has not motivated me to do my best work.

47. The instructor does not inspire class confidence by his knowledge of the subject material.

48. The instructor has given me new viewpoints and appreciations.

49. The instructor is not clear and understandable in his explanations.

50. The instructor is better than most instructors I have had.

APPENDIX B.

NUMBER OF CASES FOR CORRELATIONS IN TABLES 13, 17 AND 18

Table 22. Number of cases corresponding to the correlations for the overall group in Table 13

	LGTOTH	MAJOR	LGHSR	GPA	MSAT	ACT	HSSCI	CONFID	PLECT	PALTT	HS	FF
LGTOTHRS												
MAJOR	300											
LGHSR	272	272										
GPA	292	292	269									
MSAT	240	240	239	240								
ACT	233	233	226	231	217							
HSSCI	274	274	272	271	239	227						
CONFID	278	278	253	270	226	224	255					
PLECT	278	278	253	270	225	223	255	277				
PALTT	278	278	253	270	225	223	255	277	277			
HS	277	277	253	269	225	223	255	276	276	276		
FF	278	278	253	270	225	223	255	277	277	277	276	
RAM	276	276	252	268	224	222	254	275	276	275	276	276
TAQ	278	278	253	270	225	223	255	277	277	277	276	277
IE	272	272	247	264	220	219	249	272	271	271	270	271
LGIECK	275	275	250	267	223	221	252	275	274	274	273	274
LGIEAC	275	275	250	267	223	221	252	275	274	274	273	274
LGHRPQ	268	268	244	260	217	213	246	256	255	255	254	255
TEXT%	267	267	243	259	216	212	245	255	254	254	253	254
SGUID%	269	269	245	261	218	214	247	257	256	256	255	256
F1	266	266	242	258	215	212	244	255	254	254	253	254
F2	267	267	243	259	216	213	245	256	255	255	254	255
INSTR	263	263	239	255	214	210	241	253	252	252	251	252
SCORE	300	300	272	292	240	233	274	278	278	278	277	278

IE	LGIELK	LGIEAC	LGHRPQ	TEXT%	SGUID%	F1	F2	INSTR	SCORE
271									
272	272								
250	253	253							
249	252	252	266						
251	254	254	268	267					
249	252	252	265	264	266				
250	253	253	266	265	267	266			
248	250	251	262	261	263	262	263		
272	275	275	268	267	269	266	267	263	

Table 23. Number of cases corresponding to the correlations for the PAS section in Table 17 (males above the diagonal, females below the diagonal)

	LGTOTH	MAJOR	LGHSR	GPA	MSAT	ACT	HSSCI	CONFID	PLECT	PALTT	HS	FF
LGTOTHRS		36	32	34	27	27	32	33	32	33	33	33
MAJOR	108		32	34	27	27	32	33	32	33	33	33
LGHSR	95	95		32	27	26	32	29	28	29	29	29
GPA	105	105	94		27	26	32	31	30	31	31	31
MSAT	82	82	82	82		25	27	25	24	25	25	25
ACT	80	80	76	80	73		26	26	25	26	26	26
HSSCI	97	97	95	96	82	77		29	28	29	29	29
CONFID	106	106	94	103	81	80	96		32	33	33	33
PLECT	106	106	94	103	81	80	96	106		32	32	32
PALTT	105	105	93	102	80	79	95	105	105		33	33
HS	105	105	94	102	81	80	96	105	105	104		33
FF	106	106	94	103	81	80	96	106	106	105	105	
RAM	105	105	94	102	81	80	96	105	105	104	105	105
TAQ	106	106	94	103	81	80	96	106	106	105	105	106
IE	104	104	92	101	79	79	94	104	104	103	103	104
LGIECK	104	104	92	101	79	78	94	104	104	103	103	104
LGIEAC	106	106	94	103	81	80	96	106	106	105	105	106
LGHRPQ	102	102	90	99	77	75	92	100	100	99	99	100
TEXT%	100	100	88	97	75	73	90	98	98	97	97	98
SGUID%	102	102	90	99	77	75	92	100	100	99	99	100
F1	101	101	89	98	76	75	91	100	100	99	99	100
F2	101	101	89	98	76	75	91	100	100	99	99	100
INSTR	101	101	89	98	76	75	91	100	100	99	99	100
SCORE	108	108	95	105	82	80	97	106	106	105	105	106

RAM	TAQ	IE	LGIELK	LGIEAC	LGHRPQ	TEXT%	SGUID%	F1	F2	INSTR	SCORE
33	33	33	33	33	33	33	33	31	32	32	36
33	33	33	33	33	33	33	33	31	32	32	36
29	29	29	29	29	29	29	29	27	28	28	32
31	31	31	31	31	31	31	31	29	30	30	34
25	25	25	25	25	26	26	26	24	25	25	27
26	26	26	26	26	26	26	26	24	25	25	27
29	29	29	29	29	29	29	29	27	28	28	32
33	33	33	33	33	30	30	30	28	29	29	33
32	32	32	32	32	29	29	29	27	28	28	32
33	33	33	33	33	30	30	30	28	29	29	33
33	33	33	33	33	30	30	30	28	29	29	33
33	33	33	33	33	30	30	30	28	29	29	33
	33	33	33	33	30	30	30	28	29	29	33
105		33	33	33	30	30	30	28	29	29	33
103	104		33	33	30	30	30	28	29	29	33
103	104	103		33	30	30	30	28	29	29	33
105	106	104	104		30	30	30	28	29	29	33
99	100	98	98	100		33	33	31	32	32	33
97	98	96	96	98	100		33	31	32	32	33
99	100	98	98	100	102	100		31	32	32	33
99	100	98	98	100	101	99	101		31	31	31
99	100	98	98	100	101	99	101	101		32	32
99	100	98	98	100	101	00	101	101	101		32
105	106	104	104	106	102	100	102	101	101	101	

Table 24. Number of cases corresponding to the correlations for the TRAD section in Table 18 (males above the diagonal, females below the diagonal)

	LGTOTH	MAJOR	LGHSR	GPA	MSAT	ACT	HSSCI	CONFID	PLECT	PALTT	HS	FF
LGTOTHRS		47	42	47	40	36	42	35	35	35	34	35
MAJOR	109		42	47	40	36	42	35	35	35	34	35
LGHSR	103	103		42	39	35	42	32	32	32	31	32
GPA	106	106	101		40	36	42	35	35	35	34	35
MSAT	91	91	91	91		34	39	32	32	32	31	32
ACT	90	90	89	89	85		35	31	31	31	30	31
HSSCI	103	103	103	101	91	89		32	32	32	31	32
CONFID	104	104	98	101	88	87	98		35	35	34	35
PLECT	105	105	99	102	88	87	99	104		35	34	35
PALTT	105	105	99	102	88	87	99	104	105		34	35
HS	105	105	99	102	88	87	99	104	105	105		34
FF	104	104	98	101	87	86	98	103	104	104	104	
RAM	104	104	98	101	87	86	98	103	104	104	104	104
TAQ	105	105	99	102	88	87	99	104	105	105	105	104
IE	100	100	94	97	84	83	94	100	100	100	100	99
LGIECK	103	103	97	100	87	86	97	103	103	103	103	102
LGIEAC	101	101	95	98	85	84	95	101	101	101	101	100
LGHRPQ	99	99	94	96	84	84	94	96	96	96	96	95
TEXT%	99	99	94	96	84	84	94	96	96	96	96	95
SGUID%	99	99	94	96	84	84	94	96	96	96	96	95
F1	99	99	94	96	84	84	94	96	96	96	96	95
F2	99	99	94	96	84	84	94	96	96	96	96	95
INSTR	97	97	92	94	83	83	92	94	94	94	94	93
SCORE	109	109	103	106	91	90	103	104	105	105	105	104

RAM	TAQ	IE	LGIELK	LGIEAC	LGHRPQ	TEXT%	SGUID%	F1	F2	INSTR	SCORE
34	34	35	35	35	34	35	35	35	35	33	47
34	34	35	35	35	34	35	35	35	35	33	47
31	31	32	32	32	31	32	32	32	32	30	42
34	34	35	35	35	34	35	35	35	35	33	47
31	31	32	32	32	30	31	31	31	31	30	40
30	30	31	31	31	28	29	29	29	29	27	36
31	31	32	32	32	31	32	32	32	32	30	42
34	34	35	35	35	30	31	31	31	31	30	35
34	34	35	35	35	30	31	31	31	31	30	35
34	34	35	35	35	30	31	31	31	31	30	35
34	33	34	34	34	29	30	30	30	30	29	34
34	34	35	35	35	30	31	31	31	31	30	35
	33	34	34	34	29	30	30	30	30	29	34
104		34	34	34	29	30	30	30	30	29	34
99	100		35	35	30	31	31	31	31	30	35
102	103	100		35	30	31	31	31	31	30	35
100	101	100	100		30	31	31	31	31	30	35
95	96	92	95	93		34	34	34	34	32	34
95	96	92	95	93	99		35	35	35	33	35
95	96	92	95	93	99	99		35	35	33	35
95	96	92	95	93	99	99	99		35	33	35
95	96	92	95	93	99	99	99	99		33	35
93	94	91	93	92	97	97	97	97	97		33
104	105	100	103	101	99	99	99	99	99	97	

APPENDIX C.

GLOSSARY OF VARIABLES

Variable Name	Type[1]/Source[2]	Description
ACT	E/Arch	Composite score on American College Testing Program entrance exam.
CONFID	E/Pre	Confidence; interest in zoology, expectancy of success; four items from factor analysis.
F1	O/Post	Factor 1; perceived increase in interest in zoology, general positive evaluation of course, sense of mastery; nine items from factor analysis.
F2	O/Post	Factor 2; negative attitude toward testing and grading procedures; perceived difficulty, sense of unfairness; eight items from factor analysis.
FF	E/Pre	Fear of Failure; subscale of RAM; four items.
GENDER	E/Arch	Sex of student.
GPA	E/Arch	Cumulative grade point average prior to Spring Quarter, 1977.
HRSPERQ	P/Post	Hours per quarter of reported outside study time.
HS	E/Pre	Hope for Success; subscale of RAM; four items.
HSR	E/Arch	High school rank; percentile score; higher score = lower standing.
HSSCI	E/Arch	High school science; number of semesters of biology, physics, chemistry.

[1]Type of variable where E = entry, P = process or study pattern, and O = course outcome.

[2]Source of variable where Arch = archival data source, Pre = pre-questionnaire, and Post = post-questionnaire.

IE	E/Pre	Internal-External Locus of Control; 23 item scale; higher score = more external orientation.
IEACAD	E/Pre	IE academic; four-item subscale of IE from factor analysis.
IELUK	E/Pre	IE luck; luck versus ability as determinant of success; five-item subscale of IE from factor analysis.
INSTR	O/Post	Instructor; positive student evaluation of instructor; 11 items.
LGHRPQ	O/Post	Logarithm of HRSPERQ to effect more bell-shaped distribution.
LGIEAC	E/Pre	Logarithm of IEACAD to effect more bell-shaped distribution.
LGIELK	E/Pre	Logarithm of IELUK to effect more bell-shaped distribution.
LGTOTH	E/Arch	Logarithm of TOTHR to effect more bell-shaped distribution.
MAJOR	E/Arch	Dichotomy of student's declared major (0 = nonscience, 1 = science).
MSAT	E/Arch	Minnesota Scholastic Aptitude Test; entrance exam.
PALTT	E/Pre	Preference for alternative testing procedures; two items from factor analysis.
PAS		Phase Achievement System; one of the two instructional methods.
PLECT	E/Pre	Preference for lecture versus small group instruction; two items from factor analysis.
Post-Q		Post-questionnaire given during last week of class.

Pre-Q		Pre-questionnaire given during third lecture period.
RAM	E/Pre	Resultant achievement motivation; RAM = HS - FF; higher score = greater tendency to approach success versus avoiding failure.
SCORE	E/Arch	Score on comprehensive 80 item multiple choice final exam.
SECT		Instructional method; PAS or TRAD
SGUID%	P/Post	Reported percentage of completion of study guide.
TAQ	E/Pre	Test anxiety questionnaire; 21 items; higher score = more anxiety about testing.
TEXT%	P/Post	Reported percentage of completion of text assignments.
TOTHR	E/Arch	Total hours of college credit prior to quarter; includes transferred credit.
TRAD		Traditional instruction; one of the two instructional methods.