INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

University Microfilms International
300 North Zeib Road
Ann Arbor, Michigan 48106 USA
St. John's Road, Tyler's Green
High Wycombe, Bucks, England HP10 8HR
SOUTHWICK, Richard Nephi, 1933-
A MULTIVARIATE STUDY OF SELECTED CORRELATES
OF COURSE EVALUATIONS BY STUDENTS.

The Ohio State University, Ph.D., 1976
Psychology, general

Xerox University Microfilms, Ann Arbor, Michigan 48106
PLEASE NOTE:

Page 97 is lacking in number only, no text is missing.

UNIVERSITY MICROFILMS
A MULTIVARIATE STUDY OF SELECTED CORRELATES
OF COURSE EVALUATIONS BY STUDENTS

DISSERTATION

Presented in Partial Fulfillment of the Requirements
for the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By
Richard Nephi Southwick, B.S., M.S.

The Ohio State University
1976

Reading Committee:

Frank M. Fletcher, Adviser
Harold B. Pepinsky
Lyle D. Schmidt

Approved By

Frank M. Fletcher
Adviser
Department of Psychology
The Ph.D. degree is earned by the recipient but would be impossible to obtain without the direction, encouragement, and support of many people. I would like to express my gratitude, respect, and indebtedness to these individuals.

Dr. Frank M. Fletcher, my adviser: I am grateful to him for his patience, encouragement, and unending support. His concern for integrity and his respect for truth set high standards of excellence. I learned from his example and his word. I am proud to be his student.

Dr. Lyle D. Schmidt: As a teacher, counselor, and sounding board for me he has deeply influenced both my personal and professional development. I shall always be grateful for this.

Dr. Harold B. Pepinsky: His unreserved support and encouragement were important and deeply appreciated. I value the impact his teaching and scholarship have had on my career.

Dr. Michael J. Donovan and Dr. Naomi M. Meara: I am grateful for their assistance with statistical and methodological issues. More important, I treasure the enduring bond of friendship we share.

The Southwicks and Cheshires: I am grateful to these two remarkable families for their love and support without which this degree would not have been possible.

Our children—Rick, Stephanie, and Mark Southwick: Their love, patience, and understanding has been important. Hopefully, the hard
work, sacrifice, and perseverance they witnessed will add a
dimension of depth and understanding that will be meaningful in
their lives.

My wife Sherrie: For the endless hours of typing, for her love,
and for her unflinching faith and belief in me and my abilities
I shall never be able to adequately express my gratitude and love.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. THE PROBLEM OF PERSPECTIVE</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>A Perspective On Evaluations</td>
<td>3</td>
</tr>
<tr>
<td>Reliability</td>
<td>7</td>
</tr>
<tr>
<td>Validity</td>
<td>7</td>
</tr>
<tr>
<td>Some Correlates of Student Evaluations</td>
<td>9</td>
</tr>
<tr>
<td>Grades, Achievement, and Grade Point Average</td>
<td>10</td>
</tr>
<tr>
<td>Class Level</td>
<td>12</td>
</tr>
<tr>
<td>Personality</td>
<td>12</td>
</tr>
<tr>
<td>Teaching Rank</td>
<td>14</td>
</tr>
<tr>
<td>Class Size</td>
<td>15</td>
</tr>
<tr>
<td>Sex</td>
<td>16</td>
</tr>
<tr>
<td>Required and Elective Courses</td>
<td>17</td>
</tr>
<tr>
<td>Time</td>
<td>17</td>
</tr>
<tr>
<td>Summary of Research Literature</td>
<td>19</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>22</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>23</td>
</tr>
</tbody>
</table>
II. METHODOLOGY

Terminology
Procedures
Sample
Setting
Instruments
Course Assessment Form
Demographic and Attitudinal Questionnaire
Analysis of Data
Design

III. RESULTS AND DISCUSSION

Introduction
Hypothesis I
Hypothesis Ia
Hypothesis Ib
Hypothesis Ic
Hypothesis II
Hypothesis IIa
Hypothesis IIb
Hypothesis IIC
Hypothesis III
Hypothesis IIIa
Hypothesis IIIb
Interaction Effects: This Course
Interaction Effects: Typical Course
Interaction Effects: Difference Scores
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Comparison of CAF Intercorrelation From Two Studies</td>
<td>33</td>
</tr>
<tr>
<td>2. Probability Values of Analysis of Covariance F-Scores For</td>
<td></td>
</tr>
<tr>
<td>the Predictor Variables of Subject Field, Course Level,</td>
<td></td>
</tr>
<tr>
<td>and Sex With CAF Measures For This Course</td>
<td>38</td>
</tr>
<tr>
<td>3. Mean Scores for This Course Subject Field With CAF Measures</td>
<td>39</td>
</tr>
<tr>
<td>4. Mean Scores of Academic Predictor Variables for Each</td>
<td></td>
</tr>
<tr>
<td>Subject Field-Course Level-Sex Combination</td>
<td>43</td>
</tr>
<tr>
<td>5. Probability Values of Analysis of Covariance F-Scores For</td>
<td></td>
</tr>
<tr>
<td>the Predictor Variables of Subject Field, Course Level,</td>
<td></td>
</tr>
<tr>
<td>and Sex With CAF Measures For the Typical Course</td>
<td>45</td>
</tr>
<tr>
<td>6. Probability Values of Analysis of Covariance F-Scores For</td>
<td></td>
</tr>
<tr>
<td>the Predictor Variables of Subject Field, Course Level,</td>
<td></td>
</tr>
<tr>
<td>and Sex With CAF Measures For Difference Scores Between This and</td>
<td></td>
</tr>
<tr>
<td>Typical Course Ratings</td>
<td>47</td>
</tr>
<tr>
<td>7. Mean Scores for Difference Between This Course Rating</td>
<td></td>
</tr>
<tr>
<td>and Typical Course Rating on CAF Measures With Subject Field</td>
<td>48</td>
</tr>
<tr>
<td>8. Mean Scores For Rating of This Course on CAF Measures With</td>
<td></td>
</tr>
<tr>
<td>Course Level</td>
<td>52</td>
</tr>
</tbody>
</table>
9. Mean Scores for Difference Between This Course Rating and Typical Course Rating on CAF Measures With Course Level Including All Advanced and Introductory Courses .. 57
10. Probability Values of Analysis of Covariance F-Scores For Interaction Effect of Predictor Variables With CAF Measures For This Course ............ 60
11. Rank Order of Mean Scores on CAF Measures For Interaction of Subject Field and Course Level For This Course ................. 61
12. Rank Order of Mean Scores on CAF Measures of This Course For Interaction of Subject Field and Sex ...... 65
13. Probability Values of Analysis of Covariance F-Scores For Interaction Effect of Predictor Variables With CAF Measures For Typical Course ............. 68
14. Rank Order of Mean Scores on CAF Measures of the Typical Course For Interaction of Subject Field and Course Level ......................... 70
15. Probability Values of Analysis of Covariance F-Scores For Interaction Effect of Predictor Variables For CAF Measures and Difference Scores Between This and Typical Course Ratings ................. 77
16. Rank Order of Mean Scores For Difference Between This Course Rating and Typical Course Rating on CAF Measures With Interaction of Subject Field and Course Level ......................... 78
17. Probability Values of Analysis of Covariance F-Scores
   For Academic Predictors With CAF Measures For This
   Course, Typical Course, and Difference Scores
   Between This and Typical Course Ratings . . . . . . . . . . 93
CHAPTER I
THE PROBLEM IN PERSPECTIVE

Introduction

This research is an investigation of certain characteristics of college students that are related to student evaluations of instruction. There is a substantial body of research concerned with this topic but the findings are not conclusive and do not indicate important principles and guidelines relative to the effects of specific student variables. The focus of this study will be with the relationship of subject field, course level, sex, and traditional academic predictors to student evaluation of courses as measured by an evaluation instrument.

At first glance one might question the relationship between counseling psychology and student evaluations of instruction and by implication the interest of a counseling psychologist in the measurement and evaluation of teaching. The primary concern of the counselor is for the individual student; for facilitating the full development and utilization of his intellectual, emotional, and physical abilities in ways and areas that are meaningful for that individual. As Bordin (1969), a counseling psychologist, has indicated, teaching at its best and in its most ambitious form has always included a concern with the individuality of the student and especially with the
intellectual development of that person. The best teachers have much in common with the counselor in terms of shared values and behaviors where students are the concern. For many psychologists the teaching process and that of counseling and psychotherapy rely heavily on the utilization of learning principles (Pepinsky and Pepinsky, 1954; Rotter 1954; and Bandura 1969). There is a close relationship between the behaviors, principles and technologies that comprise the foundations of both counseling and teaching.

An important function of the counseling psychologist is the design and conduct of research for the purpose of furthering knowledge in the field. Assessment of research findings and measurement of various student characteristics are additional research-related responsibilities of the counseling psychologist in the educational setting. This concern with the quality of student attitudes and behaviors in the traditional learning context takes the counseling psychologist to areas of inquiry far beyond the counseling center such as vocational choice and career development, motivation and learning, and behaviors specific to classroom performance such as motivation, study skills, and test-taking behavior. The work of Osipow (1968) in vocational research and Bandt, Meara, and Schmidt (1974) with approaches to the overall college and classroom adjustment of students are illustrative of the breadth of the counseling psychologist's teaching, research, and counseling endeavors.

Given the background, both theoretical and practical, and orientation of the counseling psychologist it is eminently appropriate for such persons to be examining the relationships between
various student characteristics and evaluation of instruction.

A Perspective On Evaluations

Evaluation of teaching and instruction has been an important component of the educational process for as long as there have been teachers and students. It is possible to give free run to the imagination and hear the critical comments of those listening to a Socrates or an Aristotle, the parents of students learning in the primitive setting of the one-room log school querying their offspring about progress in the McGuffy readers, mastering of the "three R's," or their deportment in front of the schoolmarm. But formal evaluation as utilized in contemporary higher education became a major issue during the late 1950s and 1960s. It was during this period that concern with student input in the evaluation process and the quality of the instruments developed to facilitate that input became a major issue.

If the volume of publications is an indicator there has been a marked proliferation of interest in the process during the past six years which shows little evidence of slackening. The one characteristic of this most recent period is an interest in evaluation from the learner's point of view, not primarily a concern for the ease or meaning of higher grades but instead a reflection of the rising spirit of consumerism in contemporary society. The formal legal action filed against the university and the instructor by the student who was not admitted to a graduate program or who believed he or she failed to learn the expected skills in a specific course is becoming less uncommon. A casual reading of the Chronicle of Higher Education or
the *Journal of Law and Education* during the past year provides evidence of this trend. The warnings expressed about "grade inflation" with the consequent failure of students having obtained high grades being accepted in their application to graduate and professional schools provides additional impetus to this facet of the consumerism movement and the increasing interest in the evaluation of teaching and instruction.

The rationale for the evaluation of instruction has been articulated in diverse forms by persons with both the experience and credentials to make such pronouncements. Dressel (1970) feels that evaluation is essential for at least three reasons: for the recognition and reward of good teaching; for the knowledge and understanding which makes possible improved instruction and learning; and it helps psychological and educational researchers understand the nature of learning. In essence he says evaluation is necessary for the improvement of the educational process. Eble (1972) is more specific. He believes that the use of student evaluations are necessary for the following reasons: chances are increased that excellence in teaching will be recognized; greater student-teacher interaction may result; the institution may consider its overall goal in light of this evaluation of teaching; provision is thus made for the only direct information about faculty teaching; and a tangible sign is displaced by the faculty of the need for student involvement in goal-setting. Eble further indicated the bitter opposition he encountered with steps to assess the evaluation programs on various campuses.
Astin and Lee (1966) surveyed 1110 institutions to obtain an empirical basis for an appraisal of evaluation practices. Their data are informative. Most evaluations are by department chairmen (85 percent) and deans (82 percent). The least used formats were student evaluations (41 percent) and classroom visits (taboo at 40 percent of the institutions). Similar findings were reported by Gustad (1961). In a later report Gustad (1967) noted trends consistent with the Astin data: decline in the use of systematic student ratings; decline in classroom visitation; greater utilization of committee evaluation; greater analysis of grade distributions; wide use of informal student opinions, and evaluations by deans and chairmen; and almost total absence of research on the validity of the instruments used.

There are arguments advanced in opposition to the use of student evaluations as a consideration in promotion, tenure, and salary decisions although they are generally couched in idealistic and often obtuse language. In a recent address Aleamoni (1976) discussed six-typical faculty concerns regarding the appropriateness of using these ratings:

1. students cannot make consistent judgements because of immaturity, lack of experience, and capriciousness. Only colleagues with excellent publication records and experience are qualified to evaluate their peer's instruction.

2. most students evaluate on the basis of popularity.

3. student rating forms are neither valid nor reliable.

4. students should not make evaluations until they have
been away from the course—and possibly the university—for several years.

5. extraneous variables affect ratings, e.g., class size, schedule, sex, major, level, time of day, etcetera.

6. grade received is highly correlated with rating of course and teacher.

Each of the arguments is effectively refuted by research cited by Costin, Greenough, and Menges (1971) and Slobin (1969). Perhaps one of the more objective arguments or cautions regarding student evaluations is expressed by Fahey (1970). He notes the following: they yield pooled reactions to dimensions built into the scales; they assume uniformity of conditions, styles, and purposes; they are shaped by statistical treatment not to be additive to single scores; they are interpretable only against ideals of which no model has a consensus, or against norms which may or may not be appropriate; they have an unknown influence on teacher-student relations; and they have a built-in predisposition to establish models to be copied.

There is legitimate and long-standing support for the use of student ratings in the evaluation of teachers and instruction. The strength of that support and the utilization of such evaluation instruments is countered through strong faculty opposition. Most of the arguments opposing student evaluations tend to be focused on opinion rather than data with the exception of questions such as Fahey cited above where the issue involves statistical assumptions and procedures. The research published by Lunney (1974) offers some comfort to those worried by capricious student attitudes toward
evaluations. He discovered that graduating seniors at one institution did not want faculty personnel decisions to be heavily weighted by their evaluations. In a different study Aleamoni and Hexner (1973) found that when students were informed that their evaluations would play an important role in the promotion-tenure-salary decisions for their instructor they rated him higher than did a matched group without the personnel information. These two studies offer evidence as to the consideration and good intentions students exhibit toward the evaluation task.

The focus of this study is on the correlates of student evaluations, i.e., what variables, characteristics, and background factors influence the ways the student perceives the course as measured by ratings on evaluation instruments. Before discussing the research on correlates of student evaluations attention is directed to the question of what precisely the instruments measure and how successfully these objectives are attained, i.e., reliability and validity.

**Reliability.** A large body of evidence on questionnaires, scales, and individual items has been reviewed by Costin et al. (1971) who state unequivocally that the internal consistency and response stability of student evaluation instruments are clearly adequate.

**Validity.** In general, evaluation instruments which demonstrate adequate reliability and face validity have been used without additional validity research. The continued widespread use by students and faculty would appear to indicate their agreement that items are appropriate and that they measure what they are supposed to measure. Menges (1973) points out that validity questions are difficult to research but indicated the presence of a small body of findings
pertinent to three criteria: ratings by trained observers, end-of-course student satisfaction and learning, and alumni recollections (a possible ultimate criterion).

In a study using trained observers (Touq et al., 1973), 18 instructors were rated by their students. Two trained observers visited the classrooms and recorded the interaction data. There was considerable agreement between students and observers on what happened in class.

Ratings by faculty peers not based on classroom observations are not adequate substitutes for student ratings. According to a study by Aleamoni and Yimer (1973) the academic rank of the teacher influences colleagues' ratings. In a study done by Wilson and others (1973) students and faculty were in agreement on their nominations of "best" and "worst" teachers but disagreed on the characteristics of good teaching. Finally, along these same lines, Menges (1973) indicates that publications do not serve as a proxy for evidence of teaching effectiveness.

Regarding the intermediate criterion of student satisfaction there is significant validity evidence (Costin, et al. 1971). A more promising criterion is student learning. In 1972 research findings were published that reported a strong negative relationship between student learning and ratings of their instructors (Rodin and Rodin, 1972). The ensuing publicity and controversy ended with the publication of letters analyzing the striking weaknesses of the Rodins' study and findings from two comparable and better designed studies which found high positive relationships between learning and rating of the instructors (Gessner, 1973; Frey, 1973). A number of
researchers (Centra, 1972; Marsh, 1975) pointed out the glaring inadequacies in the Rodins' study, viz., the instructors were not evaluated; the teaching assistants who played a minor role in the actual instruction were rated. In addition there was an absence of adequate measures of performance including no final examination in the courses. Clearly a positive rather than a negative relationship is documented by the research evidence.

One interesting measure of the ultimate criterion of teacher effectiveness is student performance after college. It is extremely difficult to specify and measure the exact instructional variables that affect career achievement or emotional stability. However several studies have asked alumni to rate teachers who had positive (or negative) impacts on them. Centra (1973) found that such nominations agree with student ratings. Essentially the same findings were reported by Drucker and Remmers (1950).

The above paragraphs furnish evidence in support of the validity of student evaluation instruments, i.e., they indeed measure what they purport to measure, and that there is considerably more than face validity to these instruments.

Some Correlates of Student Evaluations

The following section reviews research findings that have some bearing on the thrust of this study, i.e., the impact of certain student and course characteristics on student evaluations of their courses. Any consideration of student evaluations of faculty and courses must include discussions about the relationship of these evaluations to variables such as grades and achievement; to class
level; to student and teacher personality, instructor rank, and class size. In addition some issues of secondary importance are included. These include the time of day during which the course is taught, of required versus elective courses, and the timing of evaluations during the semester or quarter.

Grades, Achievement, and Grade-Point Average. Anikeeff (1953) studied the effect of grading leniency on evaluation scores and found that grades reflected the ratings given by students, especially those at the freshman–sophomore level. Similar findings were reported by Kennedy (1972). His study of freshmen students in English courses found significant correlations between evaluation scores and both GPA and the grade received in the course. Weaver (1960) reported that student ratings were biased in the direction of the grades which they received in the course. These studies support the views of those who criticize these evaluations as simply a reflection of grades—the higher the grade the better the evaluation. Downie (1952) reported similar findings from 16,000 student evaluations of 406 teachers.

There is however a substantial number of studies in support of the conclusion that no significant relationship exists between grades, achievement, or grade-point average and student evaluations. Garverick (1962) could find no significant relationship between teacher ratings and grades expected or deserved by students. Similar findings were reported by Colliver (1972) and Elliott (1950). In the latter study Elliott found that teachers whose students achieved relatively more, regardless of ability, received relatively higher ratings from their students but that grades per se were not related to the evaluation
Elliott noted that some instructors are more effective with high ability students than they are with the ones with lesser ability. Rosenshine et al. (1973) reported no relationships of significance between course ratings and GPA, expected grade, and number of previous courses in the field. There was a weak tendency for students expecting higher grades to give higher ratings and for more experienced students to give higher ratings. Blum (1936), Caffrey (1969), and Voeks and French (1960) obtained results similar to the above. The latter commented that college students appear to have greater objectivity and less superficial value systems than many have realized. A significant relationship among motivation, attitude toward course, hours of study on the course, and grades received was found in a study by Mueller and Miller (1970). In a study using students from an introductory psychology course Holmes (1972) sought to determine the effects of disconfirmed grade expectancies on students' ratings of their instructor. One-half the students who deserved and expected "A's" or "B's" were given their expected grades while one-half were given a grade one step lower than expected. After receiving the grades, the students completed the course evaluation. There were no differences in evaluations as a function of differences in grades but evaluations on 11 of 19 items were lower as a function of the unexpected lowering of grades. Holmes concluded that although differences in actual grades do not affect evaluations, if students' grades disconfirm their expectancies, the students will tend to deprecate the instructor's teaching performance in areas other than his grading system.
Studies are clearly not unanimous with respect to grades, grade point average, and achievement which would suggest the need for additional research.

**Class level.** Do upper classmen rate courses and teachers higher than underclassmen do? The findings on this issue are inconclusive. Gage (1961) found that teachers of lower level courses were rated lower than the more advanced ones; Bendig (1952) indicated that upperclassmen were more unfavorable than lower classmen in their ratings; and Deutchen (1940) and Hildebrand (1971) found negligible differences in student ratings among various grade levels. Eckert and Keller (1954) and Miller (1972) found that seniors and graduate students rated courses higher than undergraduates. From these findings it would appear that one can select either position regarding the impact of course level.

**Personality.** It is not unusual to hear college teachers describe student evaluations as popularity contests or attribute the outcomes to personality variables. McKeachie (1969) concluded that students do not rate their teachers on their personality but on how they are learning. In a comprehensive study that examined many of these issues Kennedy (1972) found no correlation between student ratings and certain subscales on OPI while Ashton (1975) using a different personality measure found a significant relationship between interpersonal values of teachers as perceived by female students and their ratings of male teachers. Isaacson, McKeachie, and Milholland (1963) attempted to determine whether five personality factors generally described as relevant to teaching correlated with ratings of effective teaching.
The only high correlation (0.48) was between the peer rating of culture and student rating of effectiveness. On the basis of his research Lewis (1964) concluded that effective teachers cannot be differentiated from less effective ones on the basis of personality variables.

Guthrie (1954) was interested in the effect of showmanship and entertainment on grades. He discovered that many teachers in the top decile of the annual ratings had much in common with actors and musicians—outgoing, friendly interest in students as persons, and, most important, an industry and interest in the subject that insures his own preparation for every class period. The entertainment influence may be difficult to separate from good teaching.

A paradox of teaching was discussed by Royce (1959). That paradox is that the teacher must be popular, yet to aim at popularity is the greatest mistake a teacher can make. His study suggests that the amount of work and the standards of performance which the teacher requires are better indices of his standing in the pupil's eyes than is superficial personality attractiveness. Weaver (1960) reported that student ratings of the instructor's teaching skills did not seem to be a product of a popularity halo while Murray (1975) found the reverse to be true, that the ratings are closely related to personality traits of the instructor. He implied that the teacher's effectiveness may be determined before he enters the classroom, i.e., certain personality traits are related to classroom teaching skills which in turn are accurately rated by students. Yonge and Sassenrath (1968) found that the psychological meaning of personality factors varied from instructor to instructor, that is, there are significant differences
in correlations between the personality and rating scores from instructor to instructor, and different personality characteristics and correlated with a given factor for different instructors. The type of student who tends to rate one instructor high on a given dimension may tend to rate another instructor low on the same factor.

Many questions regarding the influence of personality variables on teaching evaluations remain unanswered by the research. It is both appealing and logical to assume that there is an impact. The issue centers on how to measure that interaction between teacher and student, and at what point do personality factors influence the evaluation.

**Teaching rank.** It is often assumed that teaching degrees and experience make a difference in the quality of teaching students receive. Research does not confirm this assumption. For example Centra (1972) found that more experienced teachers were rated no higher than those in their first or second years of teaching. Thompson (1972) reported that teaching experience had a significant impact on ratings. On the other hand Guthrie (1954) found no relationship between experience and teaching effectiveness, and slight improvement from one rank to another. Clark (1954) found that professors and associate professors were rated higher on general teaching ability than did those of lower rank. Eckert and Keller (1954) reported a direct relationship between the ratings and the rank of the instructor with professors and associate professors consistently receiving higher scores. In an early study Heilman and Armentrout (1936) found no significant differences in ratings given the same teachers over a five--to seven year period; in fact, the teachers showed a slight decrease in the latter ratings.
Downie (1952) reported no differences in evaluations between the under- and over-40 age groups, although full professors ranked highest in student-teacher relations and those with the two highest degrees had better ratings on a variety of dimensions than those with only the bachelor's degree.

It is not possible to generalize about relationships between teaching rank, and age/experience and student evaluations of teaching. It is a complex issue. Perhaps enthusiasm and a newer knowledge among younger faculty members in the lower academic ranks offsets the gains of experience and advanced degrees. Certainly the inadequacy of the research base is evident from these studies.

**Class size.** Perceptions about class size are related to the style and size of the institutions. For example, one tends to find students, faculty, and administration in small colleges favoring small classes and therefore rating them better than larger classes in the belief that small-class instruction is more personalized and therefore more effective. The above are the conclusions of Gage (1961) regarding class size. Goodhart (1948) found that classes of fewer than 20 students do not necessarily result in higher teacher ratings than those over 20. Little, if any, relationship was found by Guthrie (1954) in his study of class size and ratings given. Villano (1975) in a study involving over 14,000 student ratings of science and math classes at a large state university of 18 branch campuses did not sustain popular faculty notions ascribing highest student evaluations to so-called popular characteristics including smaller class size. Newell (1967) and Elliott (1950) both reported that rank and earned degrees make a
difference in student ratings of instruction while Colliver (1972) found no relationship between these characteristics and student ratings.

Other factors would seem to enter into the research findings such as different styles of teaching for large and small classes, the subject matter, and the tendency for college students to feel more comfortable in classes of 20 to 35 students because of 12 years of elementary and secondary school experience.

Sex. Costin and others (1971) include sex of the student and of the teacher as one of those variables not likely to significantly influence course evaluations. A number of researchers have investigated differences in ratings made by male and female students, and differences in student ratings of male and female teachers. In essence the studies report no significant differences in terms of the sex of the student or that of the teacher. As with previous issues reviewed in these pages, some evidence is to the contrary: Downie (1952) reported that female teachers were rated higher on the use of new materials; Bendig (1952) found that female students were more unfavorable than males in their ratings; McKeachie, Lin, and Mann (1971) noted that female teachers rated high in structure were rated more effective than their male counterparts. Ashton (1975) noted an interaction between female students and male teachers when the student perceived the instructor as possessing similar interpersonal values to her own. In a study cited earlier, Kennedy (1972) found that female students tended to rate their English courses higher than did males in the same courses. Finally, Centra (1972) found that female teachers were more apt to know when students did not know the material, were more concerned with student progress,
made more comments on papers or examinations, and generally made better use of class time. His other finding was that courses taught by male teachers were rated more stimulating and more difficult.

The accumulated weight of the research, especially the review by Costin and others (1971) points to an absence of significant differences in the evaluation of teaching as a function of sex, either that of the teacher or the students. Yet the evidence suggesting the presence of differences on these evaluation instruments due to sex leads to worthy issues. For example, do ratings based on sex vary according to the discipline (subject field); do sex role stereotypes affect ratings as Harris (1975) seems to indicate; do ratings based on sex vary in different types of institutions; do they reflect different teaching styles of male and female teachers?

**Required and elective courses.** Miller (1974) reports on studies of this issue and indicates that no significant differences exist between student ratings on required and elective courses. However, Gage (1961) found that teachers of elective courses received higher ratings than did instructors of required courses. Gillmore and Brandenburg (1974) reported that as the proportion of students taking a class because of a requirement increases, the ratings given to the course and instructor decrease. Again, the now-familiar pattern is evident: a general tendency in the direction of non-effect from a variable with concomitant findings to the contrary.

**Time.** Several issues are grouped under this one rubric for convenience in reporting a paucity of studies. The issues reviewed in this section are as follows: time of day the class meets, number of
times during the quarter the course is evaluated, and timing of the evaluation and instruction

Does the time of day during which the class meets make a difference in the ratings? According to Eckert and Keller (1956) classes held during the mid-portion of the day received higher ratings than early morning ones. Clark (1954) found that classes held during the middle of the day obtained higher ratings than did the early morning classes which, in turn, were regarded more favorably than were those held in late afternoon. Additional research is needed to verify or modify these findings, since the conclusions have important bearing on future course evaluation research. For example, it would be of more than casual interest to know whether those teachers lower on the academic totem pole, the young assistant professors, usually end up with the earlier morning and late afternoon classes.

A comparison of faculty evaluations early and late in the semester was conducted by Kohlan (1973) as a study of the stability of the evaluation process. The evaluation instrument was administered after the second class hour and again during the last week of the semester. No significant differences were reported between the early and late semester scores suggesting perhaps that the students' early impressions of the course are maintained throughout the semester. One is required to avoid generalizing from the findings of one limited study. However, it would appear that waiting until a course is nearly completed before collecting evaluation data might be unnecessary. Further research would be appropriate with this issue.
Summary of Research Literature. The review of prior research was restricted to the findings on eight broad issues that encompass the bulk of the investigative endeavors on student evaluations of teachers and instruction. The area of concern for this study is with the student and his perception of the learning environment and experience. For this reason some areas of research were omitted, namely findings on the impact of student evaluations on teacher behavior, such as, the influence of feedback. Of the eight correlates of student evaluations that were reviewed not one provides conclusive evidence relative to the difference its presence or absence has on the ratings. One cannot state unequivocally that female students rate courses higher or lower than males do, that higher grades are associated with better evaluations, or that smaller classes obtain higher ratings. Some research evidence is reported in support of each of these statements; other investigations detail evidence contrary to the above. Consistently throughout the literature on student evaluation of teaching one finds that additional research is needed because of the inconclusive pattern of the findings.

There are two research areas where disagreement is virtually nil: time of day during which the class meets; and impact of subject matter on the ratings. With the issue of time the class meets, the three studies reported in the literature were in agreement regarding the importance of the mid-morning hours. Subject matter of the course or subject field is a very difficult question: none of the studies reviewed dealt specifically with the impact of subject field on the ratings. There is apparently an "intuitive knowledge" or set of assumptions regarding unalterable characteristics of subject fields
that affect outcome studies. For example, within the subject field of English a course on the contemporary novel inherently captures greater interest and excitement than a course on medieval literature. This might yield inter-instructor discrepancies on ratings but not inter-subject field differences. The former pattern would have a detrimental affect on a teacher's career as a function of the kinds of ratings he obtains. Presumably each subject field is characterized by similar within field differences which offset any impact on evaluations.

The studies cited in the review of research utilized college students enrolled in courses for academic credit. The effect of the subject field has eluded the interest or attention of investigators. Goldman (1973) reported on the affect of science and non-science majors. The subjects for Mann's research were economics and history students; Kennedy (1972) used English students; Villano (1975) studied subjects in math and science courses, and Holmes (1972) utilized introductory psychology students in his research. But again the point is stressed: none of the research studies has attempted to assess the relationship between subject field and student evaluation of courses.

Both Costin et al. (1971) and Rosenshine et al. (1973) in their reviews of pertinent research findings concluded that there were no relationships of practical significance between student ratings and the variables of sex and course level among others. Despite these general conclusions there are studies providing support for the importance of both these variables: Bendig (1952, Kennedy (1972) and Ashton (1975) for the impact of sex and Gage (1961), Bendig (1952), Eckert and Keller (1954), and Miller (1972) for that of
course level.

The literature reporting on the relationships between various academic variables such as grades, grade point average, ACT and SAT scores and outcomes like success, adjustment, course evaluations, and others is extensive (Lavin 1965; and Lenning and Munday, 1974.) The findings, that is, the general pattern of the conclusions with specific reference to student evaluations are similar to those reported for course level and for sex: generally no significant relationships, with exceptions noted.

It would be of value to examine some of these variables again. The literature provides the justification for further research with virtually all variables associated with student evaluation of instruction primarily because of the dearth of conclusive findings. It is the intent of this study to examine the impact of three discrete variables, subject field, course level, and sex and their interactions on student evaluations of courses. In addition a fourth but continuous variable, academic predictors, will be included. The relationship of these variables to the student ratings of courses will be taken one step further than is normally the case, i.e., students will be asked to rate the typical course at their institution.

This latter procedure has not been attempted previously according to reviews of the research literature. Kohlan (1973) casually alluded to the possibility of rating the typical instructor in his study comparing evaluations early and late in the course and Leventhal, Lansky, and Andrews (1971) suggested that an examination of discrepancies between the ideal and actual course or teacher would be of interest. It would
appear that a study including an examination of other differences between student evaluations of their course and the typical course would be of value for several reasons: the direction of differences (if any) would provide some idea about the impact of a specific course on student perceptions of other courses; the impact of variables such as sex, subject field, course level, and grades may have on evaluations of a broader aspect of the academic experience; suggestions could be elicited relative to the student's perceptions about himself in the college environment.

The research that has been reviewed is extensive. The findings are inconclusive. It would be important to examine these correlates of student evaluation of courses: subject field, course level, sex, grades, grade point average, and SAT-verbal and math scores.

Statement of the Problem

The courses will be rated by the subjects and scores obtained on the five measures of the evaluation instrument: the Course Assessment Form (Fletcher, 1972). The Course Assessment Form is described in some detail in the following chapter. In brief it consists of 45 scaled items that provide summed scores on five factors entitled Difficulty, Content, Instruction, Examinations, and General Evaluation. It should be noted that this instrument includes five factors providing for evaluation of the total course, not just evaluation of the instructor or instruction as is the case in much of the research discussed above. The six hypotheses are based on finding differences between subject categories and characteristics on the five outcome measures listed above. When the hypothesis refers to "differences" or courses
being rated "differently" in every instance the reference is to the scores on the five Course Assessment Form measures.

Hypotheses

The following hypotheses are presented as guidelines for the research and analysis. In most cases there is not sufficient basis to attempt to predict the direction of differences therefore only differences are hypothesized.

I. It is hypothesized that the subject field will make a difference in the way students evaluate a course.
   a. Subjects will rate their specific course differently than subjects enrolled in courses in other subject fields.
   b. Subjects enrolled in a course in psychology will rate the typical course differently than students in courses in other fields, i.e., math, zoology, or education. The same hypothesis will be true for the other three academic fields, i.e., students enrolled in a specific course will evaluate the typical course differently along the dimensions indicated.
   c. There will be a difference between the ratings for the course they are enrolled in and those for the typical course as a function of the academic field of the specific course, that is, depending on whether the student is in a psychology, math, zoology, or education course.
II. It is hypothesized that course level will make a difference in the way students rate a course.
   a. Subjects enrolled in the introductory level courses will rate those courses differently than students enrolled in advanced courses.
   b. Subjects in the lower level courses will evaluate the typical course differently than subjects in the advanced courses.
   c. There will be a difference between the ratings for the course in which they are enrolled and those for the typical course as a function of the level of the course in which they are enrolled.

III. It is hypothesized that sex will make a difference in terms of how students evaluate a course.
   a. Females will rate courses differently than males.
   b. Females will rate the typical or average course differently than males.
   c. Females will differ from males in the differences between the rating scores for the course in which they are enrolled and the typical course.

IV. It is hypothesized that grade expectations will affect the students evaluation of a course.
   a. Students with higher grade expectations in their specific course will evaluate that course higher than do students with expectations of lower grades.
   b. Students with higher grade expectations will rate the
typical course more highly than students with lower grade expectations.

V. It is hypothesized that the level of performance on measures of academic potential and performance will affect students rating of a course.

a. Students with higher scores for grade point average (GPA), Scholastic Aptitude Test-verbal (SAT-V), and Scholastic Aptitude Test-mathematics (SAT-M) measures will rate their courses higher than students with lower scores.

b. Students with higher GPA, SAT-V, and SAT-M scores will evaluate the typical course in more positive directions than students with lower GPA, SAT-V, and SAT-M scores.

VI. It is hypothesized that the grade received in a course will be related to the evaluation of that course.

a. Students who obtained higher grades in their course will rate that course higher than students who receive lower grades.

b. Students who obtain higher grades in a course will rate the typical course higher than students who receive lower grades.
CHAPTER II
METHODOLOGY

Terminology

A number of words and terms are used throughout this study that require precise definition in order to avoid conflict and misunderstanding with other possible definitions.

The term subject setting is used to specify the course in which the subject is enrolled. The term course level is used to indicate the academic level of the course in which the subject is enrolled. The word measure is used to indicate a collection or set of scales which are intended to measure a single variable. The word instrument is used to indicate a collection or set of measures which are of common origin.

In this study the subject setting may be any one of four academic subjects: psychology, mathematics, zoology, or education. Course level may be either introductory or advanced. The terms lower and upper level are used interchangeably with introductory and advanced to specify course level. Introductory level courses are generally open to all students without prerequisite conditions, tend to be broad in scope, serve as a leadin to the more sophisticated courses in the department and are comprised predominantly of freshmen and sophomores. The advanced level courses require a greater degree of sophistication, assume prior experience with the content, and tend to be populated by students in their junior and senior years of college.
Procedures

Sample. The data in this study were collected at Ohio Wesleyan University in Delaware, Ohio in March of 1970 during the last week of classes prior to the formal final examination period for the Winter term. All of the 342 subjects were undergraduates variously enrolled in 15 different courses taught by 11 instructors in the departments of psychology, mathematics, zoology, and education. The courses included five sections in psychology, four in mathematics and three each from education and zoology. Of these sections seven were advanced level while eight were from the introductory category. There were 185 female and 157 male subjects. According to course level there were 194 subjects enrolled in the introductory sections and 148 in upper level courses. The organization of subjects by subject field included 115 students in psychology courses, 70 in mathematics, 97 in zoology sections and 60 in education courses.

Setting. At the beginning of the class hour on the scheduled date subjects were informed by their instructor that they were going to spend the hour evaluating the course as part of the University's institutional research and evaluation program. The writer was then introduced and proceeded with the instructions.

The subjects were directed to respond to three separate instruments with the following instructions:

1. evaluate this course using the Course Assessment Form (CAF)
2. evaluate the typical, average course at their institution, again using the CAF
3. complete a background information and attitude questionnaire (DAQ)

Assurance as to the confidentiality of names and data provided were
Several weeks after the end of the Winter term the following data were obtained from university records on each subject:

1. grade in the course
2. accumulative grade-point average through the Autumn term of 1969
3. SAT-verbal score
4. SAT-mathematics score

At the time this study was being conducted both the Scholastic Aptitude Test-verbal and Scholastic Aptitude Test-math scores were required, with few exceptions, as part of the application procedure for admission to Ohio Wesleyan University. Exceptions included applications from American students graduating from high schools outside the United States. The original sample included 378 subjects. Of this group 36 were omitted because of the absence of SAT scores or for omission of certain items on the Demographic and Attitudinal Questionnaire (DAQ).

Instruments

Course Assessment Form (CAF). Fletcher (1967, 1972) has developed an instrument on which students may report judgements about courses in which they are or have enrolled (Appendix A). The format is similar to that of the semantic differential technique developed by Osgood, Suci and Tannenbaum (1957). Rather than functioning to differentiate among several semantic stimuli, however, this instrument serves to analyze a single experience, experience in a course. The instrument employs 7-point bipolar scales with verbal judgements used to anchor each pole. The instrument yields five measurements: Difficulty, Content, Instruction, Examinations, and General Evaluations. A total of nine
scales are scored for each of the measures with a total of 45 bipolar scales. The format of the instrument is arranged so that each consecutive group of five scales includes one scale for each factor. The scales are randomly ordered within each successive group of five but so ordered that no two scales of the same factor appear next to each other. An additional feature of the CAF format is that of directionality of responses on adjacent scales. For the nine scales in each factor score four are presented in one direction and the other five are presented in the opposite direction.

The five measures or factors that comprise the CAF are described below. It is essentially the same description given by Fletcher in his 1972 publication.

I. Difficulty. The score on this measure reflects the student's perception of the relative difficulty of the course on a continuum from easy (low score) to difficult (high score). One example of a high loaded item is "Less studying than usual vs. More studying than usual."

II. Course Content. This measure could be labeled Memory vs. Thinking. The purpose is to measure the student's perception of the course content on a continuum from rote memory (low score) to emphasis on thinking or reasoning (high score). A representative scale would be "Emphasis on memory vs. Reasoning required."

III. Instruction. The student indicates his perception of the instruction including the instructor, lectures, and instruction in general on a continuum from poor (low score) to excellent (high score). A high loaded scale on this measure is usually "Poor instructor vs. Excellent instructor."
IV. **Examinations.** On a continuum from poor (low score) to good (high score) the student expresses his perception of the course examinations. Typically the course evaluation is handled prior to the final examination for practical reasons, but the timing does not appear to be crucial as long as a reasonable sample of examinations have been given in the course.

V. **General Evaluation.** This measure gives some indication of the student’s general feelings about the course on a continuum from poor (low score) to excellent (high score). The most representative scale on this factor is "Worthless course vs. Valuable course."

It is vital to the interpretation of the scores from these measures to understand subtle differences in the meaning of high and low scores for two of the factors. Content and Difficulty scores have a different meaning than scores for the other measures. For these two a high score does not necessarily have negative connotations nor is the high score automatically equated with excellence. With the Content measure a low score indicates that the emphasis in the course is perceived by the student to be upon memory and one assumes that the lower the score the greater the stress has been upon rote memorization. It is certainly possible to construe a course rated highly by the student receiving a low score on Content. This would not be unusual in science courses such as anatomy, chemistry, or many others. The same argument is applicable for the high score on Content. A course with limited requirements for memorization could offer the student little in the way of challenging intellectual fare and still obtain a high score on this measure.
Similar cautions apply to the Difficulty measure. For example, a course might require less out-of-class studying and yet be an intense, difficult course because of the demands upon the student during the class hour. The point is critical for accurate interpretation of these five measures: Instruction, Examinations, and General Evaluation scores transfer directly, i.e., high scores indicate positive feelings regarding the issue being rated; with Content and Difficulty high or low scores cannot be interpreted in the same fashion, i.e., low scores may indeed be representative of positive feelings.

A major contention of this study has been that many evaluation instruments are measures of personality and as such vulnerable to criticism by those who argue forcefully against both the reliability and validity of the data from such instruments. On the basis of its content the CAF does not appear to be a measure of personality. The instrument does not ask questions about personality nor is it so vague or ambitious that it could be liable to the accusation of being a covert measure of personality. Although it is defensible to argue that all behavior is but the manifestation of personality, it seems clear that this instrument yields responses to transient extrasubjective phenomena and is not a direct measure, either overt or covert, of basic personality characteristics.

Interpretation of the data from research on the CAF (Fletcher, 1967, 1972) both demonstrate and illustrate this point. Factor analytic techniques were employed in the development of the instrument through eight revisions. When administered to each of a number of groups of subjects, the five measures demonstrate remarkable factorial
purity and stability. If the measures were measures of personality then their intercorrelations would be stable if based on groups of subjects with comparable personalities. For example, personality characteristic A would have the same association with personality characteristic B for two sections of introductory psychology students. An examination of the intercorrelations among CAF measures for 82 classes, 56 of which were introductory psychology sections, does not show such a stable relationship between measures: the average range for the ten intercorrelations is .947. The factorial purity and stability of the measures is evidence that the variability of the intercorrelations is not attributable to unreliability on the measures. In fact the split-half reliability coefficients based on 31 sections of introductory psychology are: Difficulty, .79; Content, .73; Instruction, .90; Examinations, .90; and General Evaluation, .86. Clearly the CAF yields highly reliable measures, not of personality, but of experience in a particular course.

As part of the data analysis for this study intercorrelations were computed between the five CAF measures. The results are presented in Table 1 along with similar data reported by Donovan (1976). The intercorrelations in the two studies are quite similar. In this research the measures are moderately interrelated with the exception of Difficulty which is not related to Examinations and only modestly related to Content, Instruction, and General Evaluation.
Table 1

Comparison of CAF Intercorrelation From Two Studies

<table>
<thead>
<tr>
<th>CAF</th>
<th>Southwick</th>
<th></th>
<th></th>
<th></th>
<th>Donovan</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Difficulty</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td>.25</td>
<td></td>
<td></td>
<td></td>
<td>.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instruction</td>
<td>.28 .45</td>
<td></td>
<td></td>
<td></td>
<td>.30 .54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examinations</td>
<td>-.07 .33 .48</td>
<td></td>
<td></td>
<td></td>
<td>-.07 .44 .43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Evaluation</td>
<td>.15 .38 .66 .38</td>
<td></td>
<td></td>
<td></td>
<td>-.01 .47 .43 .49</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The reliability of the five CAF measures was computed using the Kuder-Richardson Formula Eight method. The evidence for the high reliability of this instrument with subjects in this study is substantial. For the total sample of 342 subjects reliability coefficients are .951 for Difficulty, .920 for Content, .969 for Instruction, .972 for Examinations, and .940 for General Evaluation. The coefficients for each course section range from .720 to .79 with the majority falling in the .900 range. Clearly these data indicate that the CAF is a solid instrument as judged by measurement criteria.

On the CAF measures utilized in this study four of the nine scales on each measure are reflected. This means, for example, that on Difficulty five scales are arranged so that the rightmost response position, response 7 is anchored by a verbal description of difficulty
and the selection of response 7 adds 7 points to the raw score on Difficulty. The other five scales are anchored on the right side by a verbal indication of lack of difficulty and the selection of response 7 adds only 1 point to the raw score on Difficulty. The effect of this balance in reflection is to offset any impact on raw score of laterality, a tendency to mark response positions on a given side of all scales. Reflected scales are used in all analyses after reflection, i.e., as if they had been presented in reverse form.

Demographic and Attitude Questionnaire (DAQ). The DAQ consists of 34 items assembled by the present writer for use in this study. They are an ad hoc collection of somewhat crude, unanalyzed statements designed to provide data on the subject's family and educational background plus information regarding expected grade in the course and perceptions of self in comparison to peers on a number of academic, ability, and motivational measures.

Some of the data obtained from the DAQ were not analyzed for the present study.

Analysis of Data

The data were collected on Digitek answer sheets and transferred to card form through the facilities of the Orientation and Testing Center of the Ohio State University. Data analyses were performed through the facilities of the Academic and Research Computer Center of the Ohio State University. The bulk of the computation was performed by FORTRAN programs using the BMDP2V Analysis of Covariance program and the BMDOZD Correlation with Transgeneration program.
**Design**

An analysis of Covariance design was employed using the following variables as main effects:

1. subject setting
2. course level
3. sex

An additional five variables were treated as covariates because they are continuous rather than discrete variables:

1. grade expected in course
2. grade received in course
3. cumulative grade point average (GPA)
4. SAT-verbal score
5. SAT-mathematics score

The outcome variables for this study are the summed scores of the five measures on the Course Assessment Form (CAF):

1. Difficulty
2. Content
3. Instruction
4. Examinations
5. General Evaluation
CHAPTER III
RESULTS AND DISCUSSION

Introduction

The focus of this research is with the impact of subject field, course level, and sex on the evaluations students make of a particular course. It is hypothesized that these variables will make a difference in the way students rate their courses, the typical course at their institution, and the differences between the ratings for the rated course in which they are enrolled and the typical course.

In this chapter the results for each of three major hypotheses (I, II, and III) are presented and discussed in order. Each hypotheses is organized into three sub-headings for presentation of the results for the major variable (subject field, etc.) with ratings for the specific course, the typical course, and the difference scores between these two ratings respectively.

An important component of this research is the study of the effects of the interaction of sex, subject field, and course level on the evaluation outcomes. These findings will be presented and discussed utilizing the format described above: the several interaction combinations with ratings for the specific course, typical course, and difference scores.
The last three hypotheses (IV, V, and VI) were less intensively studied. The results for these studies are presented and discussed under the heading of "Other Statistical Results."

**Hypothesis I**

The first hypothesis asserts that the Subject Field of the course makes a difference in the way the course is rated. The three aspects of this evaluation will be presented and discussed.

**Hypothesis Ia.** This hypothesis asserts that Subject Field will make a difference in the way in which the specific course is rated. The data presented in Table 2 clearly support this hypothesis which specified that a student in a psychology course would rate that educational experience differently than peers in other subject fields. In this study, for example, the other settings would be in zoology, math, or education. Comparison of F-ratios for the five dependent variables produced findings significant at the .0009 level with three of the measures, at the .002 level for one, and at the .01 level for another (see Table 2). Inspection of the respective measures of the outcome variables with subject field portray this finding clearly and at the same time provide support for the factorial purity of the CAF.

In general, subjects in psychology courses rated this course higher on the General Evaluation measure but failed to repeat this pattern on the other measures, for example Difficulty or Examinations where one might expect to find high scores consistent with the rating on General Evaluation (see Table 3). Instead subjects in zoology courses rated theirs as the most difficult, providing the best instruction and the most challenging examinations. One's expectations
Table 2
Probability Values of Analysis of Covariance F-Scores
For the Predictor Variables of Subject Field, Course Level, and Sex With CAF Measures For This Course

<table>
<thead>
<tr>
<th>CAF</th>
<th>SF</th>
<th>CL</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty</td>
<td>.002</td>
<td>.000</td>
<td>.389</td>
</tr>
<tr>
<td>Content</td>
<td>.010</td>
<td>.000</td>
<td>.201</td>
</tr>
<tr>
<td>Instruction</td>
<td>.000</td>
<td>.000</td>
<td>.677</td>
</tr>
<tr>
<td>Examinations</td>
<td>.000</td>
<td>.022</td>
<td>.935</td>
</tr>
<tr>
<td>General Evaluation</td>
<td>.000</td>
<td>.016</td>
<td>.548</td>
</tr>
</tbody>
</table>

Note: Underlined scores are significant at the .05 level or greater.
Table 3

Mean Scores for This Course Subject Field with CAF Measures

<table>
<thead>
<tr>
<th></th>
<th>Psychology</th>
<th>Math</th>
<th>Zoology</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty</td>
<td>41.4</td>
<td>34.1</td>
<td>44.0</td>
<td>35.6</td>
</tr>
<tr>
<td>Content</td>
<td>39.1</td>
<td>41.3</td>
<td>37.1</td>
<td>42.7</td>
</tr>
<tr>
<td>Instruction</td>
<td>42.3</td>
<td>35.2</td>
<td>50.5</td>
<td>43.7</td>
</tr>
<tr>
<td>Examinations</td>
<td>37.0</td>
<td>41.9</td>
<td>47.6</td>
<td>38.8</td>
</tr>
<tr>
<td>General Evaluation</td>
<td>49.3</td>
<td>35.5</td>
<td>47.9</td>
<td>44.4</td>
</tr>
</tbody>
</table>
about math courses are not supported by this particular data. With the exception of Content and Examinations, math students rate their courses lowest of the four possible subject settings, in this study—a finding that suggests that subjects make careful discriminations about the several facets of their courses.

The data for subjects in education courses is somewhat surprising in that stereotypes about the ease or relative absence of intellectual effort required in those courses is questionable. For example education courses were rated highest in terms of Content, i.e., students emphasized thinking and reasoning in these courses to a greater extent than did subjects in math, zoology, or psychology courses.

Zoology students rated their courses higher on Difficulty measure than subjects in other areas. However, these same subjects rated their courses lowest on the Content measure which suggests that zoology courses stress rote memory to a greater extent than other subject fields. Thus one is provided with an interesting profile of these courses by the CAF data. Zoology courses at the studied institution are quality educational experiences by a variety of criterion: they are extremely difficult courses because extraordinary time is required outside the classrooms and laboratory to learn the material; the emphasis is on rote memory rather than reasoning and the volume of material to be learned is significant. Because of these characteristics of the courses the examinations are challenging, and above all the teachers and their methods are stimulating to the extent that they enhance the students' experience in the course.
Math by comparison is not rated as a difficult course. The focus of the content is more on thinking than memory and the examinations are more highly rated than all but those in zoology courses. Nevertheless math courses are rated lower in terms of the general evaluation than other subject fields. Math students rated the lectures and instructors below average—more so than other subject fields were rated in this study. This finding offers further evidence in support of the thesis that students can and do discriminate between various criteria of instruction rather than making blanket judgements on a good-bad or high-low continuum.

Education courses are popularly believed to be easier than other courses, involving innumerable "doing" kinds of activities with minimal intellectual content or stimulation. The data seem to indicate that subjects enrolled in these courses evaluate them quite honestly, much as the subjects evaluate courses in other subject fields. On the Difficulty measure only math courses are rated lower by those subjects than are education courses. The high rating on Content would indicate that the emphasis is not on rote memory which is consistent with the non-education students' impressions of this subject field. Subjects in these courses rate the quality of instruction highly yet perceive the Examinations as less than good, a criticism of education courses that is not uncommon from those outside that particular subject field.

Careful inspection of covariate data in Table 4 provides support for some of the results with this main effect. With one exception, the mean SAT-V of the subjects varies as a function of course level. The SAT-V of subjects in upper level courses are greater than the scores of those in the lower levels of the same subject field with
the lone exception of subjects enrolled in education courses. The subjects in advanced education courses have SAT-V scores below those of subjects in the lower level courses. This would seem to suggest that better students tend not to persist in education courses. The pattern with other subject fields presents a more consistent picture. Presumably the winnowing process has left a group of juniors and seniors who have survived the lower level requirements, whose academic and career goals are more clearly defined, who are therefore motivated and achieving. Perhaps the marginal, the unsure, the less able students are present in these advanced courses to a lesser extent than they are in the introductory sections. The subject fields selected for this study (math, zoology, and psychology) are often regarded as somewhat challenging fields, at least more so than education. This is not meant nor intended to demean the legitimacy or importance of education, the very heart and technology of learning per se. However, in the undergraduate population education courses occupy a somewhat lower status ranking.

Assuming most students in education are aware of this it would be logical for them to defend, to justify their enrollment and commitment by reducing the dissonance they experience. This would be reflected by the elevation of scores on the CAF measures. The data presented in Table 3 indicates that subjects in education courses rate their courses lowest of the four subject fields on Difficulty (35.6). The perception of Instruction is that it is clearly above average (43.7) and of acceptable quality while Examinations are rated as being less than good (38.8). In general they rate the education
Table 4

Mean Scores of Academic Predictor Variables for Each

Subject Field-Course Level-Sex Combination

<table>
<thead>
<tr>
<th>SF</th>
<th>CL</th>
<th>Sex</th>
<th>GE</th>
<th>GR</th>
<th>GPA</th>
<th>SAT-V</th>
<th>SAT-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psych</td>
<td>Int</td>
<td>Male</td>
<td>2.88</td>
<td>2.80</td>
<td>2.36</td>
<td>541</td>
<td>600</td>
</tr>
<tr>
<td>Psych</td>
<td>Int</td>
<td>Female</td>
<td>2.89</td>
<td>2.68</td>
<td>2.72</td>
<td>556</td>
<td>570</td>
</tr>
<tr>
<td>Psych</td>
<td>Adv</td>
<td>Male</td>
<td>2.60</td>
<td>2.66</td>
<td>2.74</td>
<td>560</td>
<td>605</td>
</tr>
<tr>
<td>Psych</td>
<td>Adv</td>
<td>Female</td>
<td>2.70</td>
<td>2.63</td>
<td>2.83</td>
<td>589</td>
<td>609</td>
</tr>
<tr>
<td>Math</td>
<td>Int</td>
<td>Male</td>
<td>2.84</td>
<td>2.57</td>
<td>2.56</td>
<td>527</td>
<td>607</td>
</tr>
<tr>
<td>Math</td>
<td>Int</td>
<td>Female</td>
<td>2.72</td>
<td>2.77</td>
<td>2.81</td>
<td>542</td>
<td>589</td>
</tr>
<tr>
<td>Math</td>
<td>Adv</td>
<td>Male</td>
<td>3.71</td>
<td>3.57</td>
<td>3.10</td>
<td>599</td>
<td>683</td>
</tr>
<tr>
<td>Math</td>
<td>Adv</td>
<td>Female</td>
<td>3.08</td>
<td>3.58</td>
<td>2.98</td>
<td>592</td>
<td>690</td>
</tr>
<tr>
<td>Zoo</td>
<td>Int</td>
<td>Male</td>
<td>2.80</td>
<td>2.64</td>
<td>2.59</td>
<td>549</td>
<td>580</td>
</tr>
<tr>
<td>Zoo</td>
<td>Int</td>
<td>Female</td>
<td>2.54</td>
<td>2.54</td>
<td>2.68</td>
<td>579</td>
<td>581</td>
</tr>
<tr>
<td>Zoo</td>
<td>Adv</td>
<td>Male</td>
<td>2.96</td>
<td>3.03</td>
<td>3.02</td>
<td>585</td>
<td>634</td>
</tr>
<tr>
<td>Zoo</td>
<td>Adv</td>
<td>Female</td>
<td>3.00</td>
<td>2.75</td>
<td>2.86</td>
<td>594</td>
<td>614</td>
</tr>
<tr>
<td>Ed</td>
<td>Int</td>
<td>Male</td>
<td>3.10</td>
<td>3.30</td>
<td>2.22</td>
<td>528</td>
<td>578</td>
</tr>
<tr>
<td>Ed</td>
<td>Int</td>
<td>Female</td>
<td>3.00</td>
<td>3.90</td>
<td>2.53</td>
<td>556</td>
<td>563</td>
</tr>
<tr>
<td>Ed</td>
<td>Adv</td>
<td>Male</td>
<td>4.00</td>
<td>4.00</td>
<td>2.59</td>
<td>479</td>
<td>545</td>
</tr>
<tr>
<td>Ed</td>
<td>Adv</td>
<td>Female</td>
<td>3.51</td>
<td>3.55</td>
<td>2.61</td>
<td>506</td>
<td>530</td>
</tr>
</tbody>
</table>
course positively as a valuable course (44.4) but not as high as it might be rated because the value of the course stems not from the intellectual time demands nor challenging examinations but from the content of the course and the manner in which it is presented. In essence they seem to imply that the education courses are important, worthwhile but not as demanding or rigorous as other subject fields.

Hypothesis Ia is clearly supported by the data summarized in Table 2 and Table 3. The subjects apparently rate their courses objectively because the evidence would mitigate against a general blanket rating of courses in a subject field. Indeed students in psychology courses rate their courses differently than do students enrolled in courses in other subject fields. The same statement is applicable to students in math, zoology, or education courses. Perhaps somewhat surprising is the data from education courses. Contrary to popular stereotypes these courses appear to be rated as carefully and uniquely as the other courses.

Hypothesis Ib. This hypothesis asserts that the subject field of the course will make a difference in the way the typical course is rated on the five CAF measures. This hypothesis was not supported by the data. None of the F-scores approached the .05 level of significance. These data are presented in Table 5. One must conclude, therefore, that although subject field has a clear and definite relationship to the way in which the specific course is evaluated, it does not affect this evaluation of the average or typical course. The task at hand when evaluating "this" course is obvious, the relevance of each scale is direct and the subjects presumably recall specific
Table 5

Probability Values of Analysis of Covariance F-Scores
For the Predictor Variables of Subject Field, Course Level, and Sex With CAF Measures For the Typical Course.

<table>
<thead>
<tr>
<th>CAF</th>
<th>Typical Course</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SF</td>
<td>CL</td>
<td>Sex</td>
</tr>
<tr>
<td>Difficulty</td>
<td>.226</td>
<td>.003</td>
<td>.007</td>
</tr>
<tr>
<td>Content</td>
<td>.223</td>
<td>.567</td>
<td>.898</td>
</tr>
<tr>
<td>Instruction</td>
<td>.151</td>
<td>.488</td>
<td>.517</td>
</tr>
<tr>
<td>Examinations</td>
<td>.354</td>
<td>.796</td>
<td>.577</td>
</tr>
<tr>
<td>General Evaluation</td>
<td>.397</td>
<td>.445</td>
<td>.060</td>
</tr>
</tbody>
</table>

Note: Underlined scores are significant at the .05 level or greater.
behaviors, experiences, or feelings in support of the rating they make. This is apparently not true when they consider the "typical" course.

**Hypothesis Ic.** This hypothesis asserts that there will be a difference between rating scores for the specific course and those for the typical course as a function of subject field.

There is statistical evidence in support of the hypothesized difference scores between ratings of this course and those for the typical course (Table 6). The hypothesized differences on the Content, Examinations, and General Evaluation measures were supported. These data are presented in Table 7. The difference scores for Difficulty and Instruction were not statistically significant.

The earlier discussion for Hypothesis Ia continues to be applicable for the impact of subject-field on the difference scores (Hypothesis Ic). There appears to be a "not so subtle" tendency among subjects in education courses to rate their courses in an extremely positive way in comparison to the perceived typical course. Education subjects believe their courses emphasize reasoning far more than the typical course and to a greater extent than do students rating their own courses whether in psychology, math, or zoology. The suggestion here is the need on the part of subjects in education courses to genuinely believe their courses measure up to and indeed surpass other subject fields in terms of the quality and intellectual rigor of their courses.

The difference scores for subjects in math courses maintain a significant and somewhat confusing pattern. On the Content measure math courses are rated higher than the typical course indicating that
Table 6
Probability Values of Analysis of Covariance F-Scores
For the Predictor Variables of Subject Field, Course Level, and Sex With CAF Measures For Difference Scores Between This and Typical Course Ratings.

<table>
<thead>
<tr>
<th>CAF</th>
<th>Difference Score</th>
<th>SF</th>
<th>CL</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty</td>
<td>.184</td>
<td>.000</td>
<td>.334</td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td>.030</td>
<td>.000</td>
<td>.369</td>
<td></td>
</tr>
<tr>
<td>Instruction</td>
<td>.140</td>
<td>.001</td>
<td>.402</td>
<td></td>
</tr>
<tr>
<td>Examinations</td>
<td>.005</td>
<td>.024</td>
<td>.594</td>
<td></td>
</tr>
<tr>
<td>General Evaluation</td>
<td>.000</td>
<td>.013</td>
<td>.347</td>
<td></td>
</tr>
</tbody>
</table>

Note: Underlined scores are significant at the .05 level or greater.
Table 7

Mean Scores for Difference Between This Course Rating and Typical Course Rating on CAF Measures with Subject Field

<table>
<thead>
<tr>
<th>CAF</th>
<th>Psychology</th>
<th>Math</th>
<th>Zoology</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>4.3</td>
<td>3.7</td>
<td>-0.9</td>
<td>6.4</td>
</tr>
<tr>
<td>Examinations</td>
<td>-0.4</td>
<td>2.7</td>
<td>7.8</td>
<td>3.5</td>
</tr>
<tr>
<td>General Evaluation</td>
<td>9.6</td>
<td>-4.6</td>
<td>6.4</td>
<td>4.8</td>
</tr>
</tbody>
</table>

Note: The negative (-) sign indicates a difference score in favor of the typical course on the particular CAF measure.
these subjects perceive the typical course as requiring greater emphasis on rote memory than math courses. Nevertheless, subjects in both psychology and education courses see these differences as being greater than those in math courses. One might infer from these differences that subjects at this particular institution rate their fields of study favorably and take significant pride in the calibre of instruction.

Attention is directed to the score for math on the general evaluation measure. The difference score of 4.6 in favor of the typical course requires explanation. Subjects in math courses appear to discriminate between the five measures when evaluating these courses. They differentiate between course Content, Examinations, and General Evaluation of their courses and their perception of the same aspects of the typical course quite specifically. For example, both Content and Examinations are significantly better, more challenging in their specific courses than in the typical courses. But when they make judgements about the overall quality of their courses other variables, other issues have an important bearing on the ratings. Although the differences were not statistically significant one might suggest that difficulty or feelings about the instructor might affect the overall rating. Certainly students in math courses have positive feelings about the typical course in contrast to their math courses.

There appears to be an important selection factor operating to explain the ratings in education and math courses but in quite different ways. The stereotype of the college math course is one of extraordinary difficulty for non-majors. The usual introductory course in math is calculus, a course that reinforces the stereotype. Many
students self-select themselves out of the introductory math pool so that the math sample included in this study has already been restricted. The process continues for students who remain in math or continue to take courses through the advanced levels. The process for students in education courses is similar but the direction is reversed: no decision to avoid education because of the intellectual challenge involved; instead a movement out of education courses for many reasons including the absence of structure and intellectual challenge among others.

As has been characteristic of other subject fields, students in zoology courses respond to the assignment of evaluating courses with a degree of objectivity. This seems to be demonstrated by a pattern of discriminating among the variables being evaluated. Reference is again made to Table 4. The difference score for these subjects on the Content measure points out that they see relatively minimal differences between their zoology courses and the typical course. If anything, they indicate a belief that the typical course may place a stronger emphasis on thinking and reasoning than zoology courses. The difference score on Examinations suggest that these students believe examinations are much better, i.e., difficult, rigorous, demanding, in zoology courses than in the typical course. The magnitude of that difference surpasses the other subject fields on this dimension. In a similar fashion the General Evaluation of zoology courses reflects the positive ratings subjects gave those courses in comparison to the manner in which they rated the typical course on the same measure.

The importance of the perceived difference between the course in which the subject is enrolled and the typical course is evident
from the data. It is not simply that subjects evaluate the two course settings differently: it is the direction and magnitude of the difference that is crucial. Subjects do rate the two settings differently but the importance of the difference can be seen by noting the difference scores for the General Evaluation measure: the direction is the same, i.e., in favor of this course, but the magnitude of that difference between psychology and education subjects communicates the subjects' feelings about the particular subject field (psychology) in comparison to the feelings of subjects in other courses (education) and in comparison to the typical course. A similar statement could be made regarding math. Clearly subjects in those courses view it as significantly different from the typical course and rate math courses in a fashion at odds with the manner subjects in the other subject fields rate their courses.

**Hypothesis II**

The second hypotheses asserts that the course level of the course in which the subjects are enrolled will make a difference in the way the course is rated. The three aspects of this evaluation will be discussed in order.

**Hypothesis IIa.** This segment of hypothesis II asserts that course level will make a difference in the way in which the particular course is evaluated. The hypothesis is supported by the data presented in Table 2. The F-scores for course level with the five CAF variables are significant at levels ranging from .022 for Examinations to .0009 for Difficulty. The mean scores are presented in Table 8. On all but one of the measures subjects in the advanced level rated their
Table 8
Mean Scores For Rating of This Course on CAF Measures With Course Level

<table>
<thead>
<tr>
<th>CAF Measures</th>
<th>Advanced Courses</th>
<th>Introductory Courses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty</td>
<td>42.7</td>
<td>37.5</td>
</tr>
<tr>
<td>Content</td>
<td>42.6</td>
<td>37.4</td>
</tr>
<tr>
<td>Instruction</td>
<td>46.2</td>
<td>41.4</td>
</tr>
<tr>
<td>Examinations</td>
<td>39.8</td>
<td>42.6</td>
</tr>
<tr>
<td>General Evaluation</td>
<td>47.9</td>
<td>43.2</td>
</tr>
</tbody>
</table>
courses higher on the CAF than did subjects in the introductory courses. The lone exception was on Examinations where subjects enrolled in introductory level courses rated their courses higher than did subjects in the upper level courses. These results are both consistent and expected. One would logically assume that subjects in the advanced level courses would rate them higher on Content, Difficulty, and Instruction as a function of expectations generated by experience. At this level one expects to function at a level beyond the rote memory required for the more basic courses and thus expect the emphasis to be on higher cognitive skills such as thinking, reasoning, and conceptualizing constructed upon the basic foundation one has acquired in the particular subject field. Perhaps because of the level of functioning and the restricted focus more time and energy will be required in order to succeed in the advanced courses and subjects will evaluate them as more difficult than the courses at the introductory level. The score on the Instruction measure reflects the subjects' view of the teacher, the lectures, and the instructional style. The higher mean score on this measure indicates that subjects in the upper level courses perceive these characteristics from a different perspective than subjects enrolled in lower level courses, that is, they believe the advanced course tends to be in the instructor's area of specialization and thus all his knowledge and skills are more effectively utilized in the teaching context with a more select group of students to the mutual benefit of both parties.
Why then the lower rating on Examinations by subjects in the advanced course? Subjects in the lower level courses are predominantly freshmen and sophomores. The test-preparation and test-taking experience at the college level is a relatively newer experience, one filled with unexpected pressures, standards of performance, and consequent apprehension. It is assumed that there would be a normal tendency to evaluate examinations at this stage as difficult and challenging—in a word, as "good" examinations. For the subjects in the advanced courses taking examinations has become a routine component of college life. By the junior and senior years these subjects are "test wise". This does not necessarily mean that they all perform at substantial levels of excellence on examinations. Rather it implies that there are few surprises for students at this level. They would tend to rate examinations as easier and less challenging. It may further indicate a certain confidence on the part of the student in his ability to master the required material and demonstrate this knowledge on examinations.

The difference in the General Evaluation by the two groups would appear to be both a reflection and confirmation of the preceding characteristics. Obviously experience with a given subject field and general educational sophistication affect one's perception of a specific course. The course level of the specific class in which the subject is enrolled has a demonstrable effect on that person's evaluation of the course.
Hypothesis IIb. This aspect of Hypothesis II states that course level will make a difference in the way the typical course is rated on the five CAF measures. The data did not confirm the hypothesis except for rating of the typical course on the Difficulty measure where the relationship was significant at the .003 level (see Table 5). The mean score for advanced courses was 40.3, that of the introductory courses 45.0. This single finding is the reverse of the pattern for course level with the specific course ratings where the advanced courses were rated highest. It would appear from these findings that subjects with less experience in the college classroom environment may feel relatively confident of their ability to succeed in the introductory courses but experience some apprehension about the years in front of them. Part of the normal student mode is to inquire and speculate about the courses required for graduation and to calculate the best conceivable set of circumstances to facilitate success with these academic challenges. For example, one might obtain information regarding the teachers who might instruct, the year and quarter in which to take the course, and the courses to enroll in concurrently with the subject of the student’s apprehension. This information assists in the decision-making process, a process which may be construed as a necessity and stemming from the introductory level student’s concern with the academic barriers to graduation.

The general lack of significant findings between course level and the CAF measures with the typical course parallels the data reported for subject matter with the typical course. The plausible explanation for the lack of significant results is the same as that offered for the
subject field with typical course failure: lack of immediate relevance to the student when the analysis concerns a single predictor variable.

**Hypothesis IIc.** This hypothesis asserts that there will be a difference between rating scores for the specific course and those for the typical course as a function of course level. The findings support the hypothesis on all five CAF measures at levels of significance ranging from .0009 for Instruction to .030 for Content. These data are found in Table 6: the difference scores are presented in Table 9. Although the relationships were not significant when rating the typical course per se, the differences between the specific and typical ratings are meaningful in that the scores portray sharp differences in the way the two course settings are perceived on the evaluation measures. In general students in both the advanced and the introductory courses rated their courses higher on the CAF measures than the typical course. Students enrolled in the introductory courses viewed the typical course as being far more difficult than their particular courses while this same course level saw little difference between the two levels on Content and Instruction. Students in the advanced courses rated all five CAF measures higher for their courses than for the typical course. The differences were especially great on General Evaluation, Instruction, and Content.

**Hypothesis III**

Hypothesis III asserts that the sex of the subject will make a difference in the way the course is rated.

**Hypothesis IIIa.** This portion of the hypothesis states that the sex of the subject will make a difference in the way the particular
Table 9
Mean Scores for Difference Between This Course Rating and Typical Course Rating on CAF Measures with Course Level Including All Advanced And Introductory Courses

<table>
<thead>
<tr>
<th>CAF Measures</th>
<th>Advanced</th>
<th>Introductory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty</td>
<td>3.64</td>
<td>-7.60</td>
</tr>
<tr>
<td>Content</td>
<td>8.09</td>
<td>-0.78</td>
</tr>
<tr>
<td>Instruction</td>
<td>7.28</td>
<td>0.65</td>
</tr>
<tr>
<td>Examinations</td>
<td>3.29</td>
<td>3.18</td>
</tr>
<tr>
<td>General Evaluation</td>
<td>8.51</td>
<td>2.19</td>
</tr>
</tbody>
</table>

Note: Negative sign preceeding scores indicates the direction of the difference is in favor of the typical course.
Hypothesis IIIb. This part of Hypothesis III asserts that the sex of the subject will make a difference in the way the typical course is evaluated. This hypothesis was not supported by the data except for the relationship between sex and Difficulty which was significant at the .007 level (see Table 5). The impact of the subjects' sex on the Difficulty measure is very direct. Females rated the typical course as more difficult than did males. The mean score for females is 44.4 and for males 41.4. It is readily apparent from the data that sex of the subjects has no measureable effect on how the courses are evaluated except this one variable for the typical course. There are several possible explanations for this finding. The well-documented (O'Leary, 1974) need on the part of the college female to achieve and to avoid failure in the academic arena may facilitate an attitude toward courses in general which motivates her to study and achieve at a higher intellectual level. This attitude conceivably might be that all college courses will be difficult. If, for example, the female student finds the course she is enrolled in to be relatively easy her response might be that if she is doing well in the course it must not be too difficult and therefore less challenging than the typical courses.

Interaction Effects

The two-way interaction effects of this study are the source for some of the important findings of this research. The results will be discussed in the following order: the evaluations of the present course, that of the typical course, and last for the difference scores between the two ratings.
Interaction of Subject Field and Course Level: This Course. The effect of both subject field and course level combined on the dependent variable measures was significant at levels from .035 to .0009 (Table 10). The interaction was not significant for the Difficulty score alone among the five measures comprising the CAF.

The scales comprising the Content measure when marked at the upper extremity result in a score indicating that the course content stresses thinking and reasoning; a low score would mean the emphasis is on rote memory. One of the assumptions upon which the CAF is based is that course content stressing more abstract intellectual processes is highly valued while rote memorization activities are not. It is evident from this interaction effect that subjects in the advanced math course view their courses as emphasizing these intellectual processes to a greater extent than subjects in other SF-CL combinations. These data are presented in Table 11-A.

At the other extreme math subjects in the introductory sections rate their courses among the lowest on the Content measure (36.7). This pattern of low ratings for lower level math courses is found on the other CAF measures. Although not the lowest, the evaluation of Examinations was among the group of SF-CL combinations obtaining poor ratings (36.9). This pervasive tendency in the introductory math sections implies that a number of variables are interacting including one not measured in this study: specific teacher personality and teaching behaviors. The low rating for Content illustrates the subjects' perception of the importance placed upon rote memory in those courses. One would assume a necessity for memorization in the introductory
Table 10
Probability Values of Analysis of Covariance F-Scores
For Interaction Effect of Predictor Variables
With CAF Measures For This Course.

<table>
<thead>
<tr>
<th>CAF</th>
<th>SF-CL</th>
<th>SF-Sex</th>
<th>CL-Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty</td>
<td>.186</td>
<td>.413</td>
<td>.789</td>
</tr>
<tr>
<td>Content</td>
<td>.012</td>
<td>.044</td>
<td>.954</td>
</tr>
<tr>
<td>Instruction</td>
<td>.000</td>
<td>.876</td>
<td>.943</td>
</tr>
<tr>
<td>Examinations</td>
<td>.000</td>
<td>.869</td>
<td>.994</td>
</tr>
<tr>
<td>General Evaluation</td>
<td>.035</td>
<td>.510</td>
<td>.205</td>
</tr>
</tbody>
</table>

Note: Underlined scores are significant at the .05 level or greater.
Table 11
Rank Order of Mean Scores on CAF Measures for Interaction of Subject Field And Course Level For This Course

<table>
<thead>
<tr>
<th>A. SF-CL with Content</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math Advanced</td>
<td>54.1</td>
</tr>
<tr>
<td>Education Advanced</td>
<td>44.9</td>
</tr>
<tr>
<td>Education Introductory</td>
<td>40.7</td>
</tr>
<tr>
<td>Psychology Advanced</td>
<td>40.2</td>
</tr>
<tr>
<td>Zoology Advanced</td>
<td>38.8</td>
</tr>
<tr>
<td>Psychology Introductory</td>
<td>37.4</td>
</tr>
<tr>
<td>Math Introductory</td>
<td>36.7</td>
</tr>
<tr>
<td>Zoology Introductory</td>
<td>36.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. SF-CL with Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoology Introductory</td>
</tr>
<tr>
<td>Education Advanced</td>
</tr>
<tr>
<td>Psychology Advanced</td>
</tr>
<tr>
<td>Math Advanced</td>
</tr>
<tr>
<td>Zoology Advanced</td>
</tr>
<tr>
<td>Psychology Introductory</td>
</tr>
<tr>
<td>Education Introductory</td>
</tr>
<tr>
<td>Math Introductory</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. SF-CL with Examinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math Advanced</td>
</tr>
<tr>
<td>Zoology Introductory</td>
</tr>
<tr>
<td>Education Advanced</td>
</tr>
<tr>
<td>Psychology Introductory</td>
</tr>
<tr>
<td>Zoology Advanced</td>
</tr>
<tr>
<td>Math Introductory</td>
</tr>
<tr>
<td>Psychology Advanced</td>
</tr>
<tr>
<td>Education Introductory</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D. SF-CL with General Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education Advanced</td>
</tr>
<tr>
<td>Zoology Introductory</td>
</tr>
<tr>
<td>Psychology Advanced</td>
</tr>
<tr>
<td>Psychology Introductory</td>
</tr>
<tr>
<td>Zoology Advanced</td>
</tr>
<tr>
<td>Math Advanced</td>
</tr>
<tr>
<td>Education Introductory</td>
</tr>
<tr>
<td>Math Introductory</td>
</tr>
</tbody>
</table>
courses as part of the acquisition of the requisite foundation in the discipline. The discrepancy between the scores for advanced and introductory math subjects on Content alone seems extreme (54.1 and 36.7). The difference between the two math groups on Instruction (44.8 and 31.6) is perhaps the crucial data allowing one to conclude that factors other than those specifically measured in this study affected the ratings for the introductory level math courses.

It is important to again recognize the relative factorial purity of the CAF as demonstrated by the data on each of the five measures. The subjects discriminate between the different CAF measures and the various subtleties of each even when the general quality of the course and teacher is uniformly poor. Witness the ratings on Examinations for the introductory math courses (Table 11-C). Although the examinations were not highly regarded, subjects rated these courses higher on this specific dimension than they did on Instruction or General Evaluation.

Data from the education SF-CL interaction points out further characteristics about these courses and students enrolled in them. Scores on the measure of Content are not particularly discriminating, both levels of education courses ranking near the top of the ratings. This would indicate that subjects at both levels in this particular subject field perceive the emphasis as being on thinking and reasoning rather than rote memory. On the other hand, rather than stressing the thinking aspects of Content it may be a reflection of the absence of emphasis on rote memory that produces the high Content scores. The valid explanation may lie with the vagueness of education per se:
there is little to memorize, little of a definitive nature in many of these undergraduate courses. It may also be a function of how the courses are taught—whether the emphasis is on content or process—and this research may have sampled education courses whose teachers focus on process.

The interaction effect is obvious with the discrepancy between the two course level groups on the Instruction and Examination measures (Tables 11-B and 11-C). Subjects in the advanced sections rated the quality of instruction very high (52.7), among the highest of any group on this measure while subjects rated the introductory level significantly below average (35.9) and near the lowest for any SF-CL grouping. It would appear that the evaluations are not solely a function of the SF or CL but of the interaction of these two variables. A further demonstration of this is observed in the scores for zoology courses. The moderate and low rating for advanced and introductory sections respectively reflects the importance of the memorization task at both levels, but especially at the introductory level. One might assume, as was the case when examining the main effects, that advanced CL sections would be rated higher on Instruction again because of the backlog of experience of the subjects in the SF and the expertise and motivation of the teacher. Subjects in the lower level courses rated their course extremely high on Instruction whereas subjects in the advanced section were quite moderate in evaluating that aspect of the course. The same pattern is apparent in the way they evaluated Examinations and with the General Evaluation. In both settings subjects rated the introductory sections highly (50.4 and 49.7) and advanced level
significantly lower (41.4 and 44.2).

Further evidence of the sensitivity of both subjects and the instrument to the interaction effects can be observed in the data from psychology courses. Whether subjects rate the advanced or introductory courses more favorably is a function of the variable assessed. For example on Content, subjects in the advanced sections acknowledge a stronger emphasis upon more abstract intellectual processes while those on the introductory courses perceive the content as clearly memorization oriented. While the quality of instruction was rated higher by those in advanced courses, subjects in lower level psychology courses felt the examinations were of higher quality than did subjects in advanced sections. The results of the General Evaluation find the subjects at both course levels rating their sections favorably to the same degree numerically. Thus the subjects do not simply give a broad general evaluation of their particular course but appear to respond carefully to the salient aspects of the specific experience they are directed to evaluate.

Subject Field and Sex with Content: This Course. The interaction of Subject Field (SF) and Sex was statistically meaningful only with the dependent variable of Content for the specific course (Table 12). The absence of significant findings on this interaction is a reflection of the pervasive absence of any effects of sex upon the hypothesized results. One might conclude from this isolated finding that the sex of the subject affects how course content is perceived. Sex makes little difference in the way subjects in education courses rate the content of their courses. Both male and female subjects rate their
Table 12

Rank Order of Mean Scores on CAF Measures of This Course
For Interaction of Subject Field and Sex

<table>
<thead>
<tr>
<th>SF-Sex With CAF Content</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math Females</td>
<td>46.2</td>
</tr>
<tr>
<td>Education Females</td>
<td>42.8</td>
</tr>
<tr>
<td>Education Males</td>
<td>42.5</td>
</tr>
<tr>
<td>Psychology Males</td>
<td>39.4</td>
</tr>
<tr>
<td>Zoology Females</td>
<td>39.1</td>
</tr>
<tr>
<td>Psychology Females</td>
<td>38.9</td>
</tr>
<tr>
<td>Math Males</td>
<td>37.8</td>
</tr>
<tr>
<td>Zoology Males</td>
<td>35.4</td>
</tr>
</tbody>
</table>
sections high on the Content dimension indication that these students perceive education courses placing strong emphasis upon the more abstract and intellectual components of the subject matter. Contrast this finding with the data from math courses. The interaction of SF with sex has dramatic results. Women in math courses rate their courses higher on the Content measure than all other SF-Sex combinations: males are near the opposite end of the continuum. Thus females in the same math courses with male subjects rated the course content high on thinking and reasoning while males evaluated the same courses low indicating that for them rote memory was the dominant characteristic of the courses in terms of the Content measure.

The examples of math and education provide evidence of the importance of the interaction effects of these two variables on the outcome variables of the CAF. The effect of sex alone is not significant nor is SF, but taken in combination the results are evident and significant.

Females in zoology courses respond differently to the Content measure than males. The former rate their courses somewhat higher on Content than do the males who see zoology courses stressing rote memory to a larger degree than other SF-Sex combinations. Interestingly, sex seems to make little difference in the ratings by subjects in psychology courses, i.e., both male and female subjects rated psychology courses roughly the same way on Content, perceiving a mild emphasis on memory with lesser concern for thinking and reasoning.

Attempts to predict how subjects will evaluate a course are enhanced by the analysis of the manner in which the variables interact.
From the preceding discussion it is obvious that the combination of the two variables (SF and CL) facilitate a more precise understanding of student evaluations. The data from SF or CL alone, although statistically meaningful, do not provide the kind of data that, for example, SF interacting with sex reveals concerning the content of a specific course. Obviously the sex of the subject and the course in which he or she is enrolled affect that particular evaluation in delicate ways. It is apparent that females in math courses respond quite differently to the task of rating that course than do females in psychology courses. One cannot make broad statements about female students or students in a specific subject field. The interaction effect is crucial to a full understanding of the evaluation of a course.

**Interaction Effects: Subject Field-Course Level, Typical Course**

As would be expected from the data based upon evaluation of courses in which the subjects are currently enrolled, the examination of findings from ratings of the typical courses are significant only for the SF-CL interaction (Table 13). These interactions are significant with all five of the CAF measures—Difficulty, Content, Instruction, Examinations, and General Evaluation. These results include all the interactions found with the evaluations of the specific course in addition to significant results for this interaction on the Difficulty measure. A cursory examination of the two sets of data (Table 11 and Table 14) leads to several general conclusions: (1) a tendency to rate specific courses higher than typical courses, (2) wider range between high and low scores on each of the CAF with the specific course, measures, (3) different rank order for Subject Fields under
Table 13

Probability Values of Analysis of Covariance F-Scores
For Interaction Effect of Predictor Variables
With CAF Measures For Typical Course.

<table>
<thead>
<tr>
<th>CAF</th>
<th>SF-CL</th>
<th>SF-Sex</th>
<th>CL-Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty</td>
<td>.001</td>
<td>.731</td>
<td>.068</td>
</tr>
<tr>
<td>Content</td>
<td>.001</td>
<td>.123</td>
<td>.557</td>
</tr>
<tr>
<td>Instruction</td>
<td>.021</td>
<td>.502</td>
<td>.684</td>
</tr>
<tr>
<td>Examinations</td>
<td>.003</td>
<td>.701</td>
<td>.782</td>
</tr>
<tr>
<td>General Evaluation</td>
<td>.030</td>
<td>.627</td>
<td>.549</td>
</tr>
</tbody>
</table>

Note: Underlined scores are significant at the .05 level or greater.
the two separate rating conditions (this course vs typical course), and
(4) the tendency for subjects enrolled in the introductory level psych-
ology courses to rate typical courses highest and for subjects in
advanced psychology courses to rate the typical course lowest on the
CAF measures. The full impact of these findings becomes more apparent
as the specific interactions are discussed.

Subject Field and Course Level with Difficulty: Typical Course.
Patterns similar to those found with evaluation of the course in which
the subjects were enrolled are observed when the same subjects evaluate
the typical course (Table 14). The combination of SF and CL determines
the perception of the typical course as indicated by the magnitude of
the scores. Subjects in different Subject Fields at different Course
Levels rate the typical course quite uniquely. Subjects in the lower
level psychology courses rank the typical course highest in terms of
content (46.9) while subjects in the advanced psychology courses rate
it lowest on the same measure (38.1). One might predict this result
as a function of Course Level. The introductory student tends to
believe that by definition the more advanced courses will be more
demanding; the wise, sophisticated junior or senior student "knows"
his particular course is difficult and to a greater extent than the
typical course. This explanation does not, however, shed light on
the impact of Subject Field (psychology in this instance) on these
findings. If this were the adequate explanation then a similar pattern
would be observed for the other Subject Fields; education, math, and
zoology. However, subjects in education courses rate the typical
course in a similar vein regardless of course level as do subjects
Table 14

Rank Order of Mean Scores on CAF Measures of the Typical Course
For Interaction of Subject Field and Course Level

<table>
<thead>
<tr>
<th>A. SF-CL with Difficulty</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychology Introductory</td>
<td>46.9</td>
</tr>
<tr>
<td>Education Introductory</td>
<td>45.9</td>
</tr>
<tr>
<td>Education Advanced</td>
<td>44.7</td>
</tr>
<tr>
<td>Math Introductory</td>
<td>44.2</td>
</tr>
<tr>
<td>Zoology Introductory</td>
<td>44.0</td>
</tr>
<tr>
<td>Zoology Advanced</td>
<td>42.5</td>
</tr>
<tr>
<td>Math Advanced</td>
<td>39.1</td>
</tr>
<tr>
<td>Psychology Advanced</td>
<td>38.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. SF-CL with Content</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychology Introductory</td>
<td>39.8</td>
</tr>
<tr>
<td>Math Introductory</td>
<td>39.2</td>
</tr>
<tr>
<td>Zoology Advanced</td>
<td>38.4</td>
</tr>
<tr>
<td>Education Advanced</td>
<td>38.0</td>
</tr>
<tr>
<td>Zoology Introductory</td>
<td>39.9</td>
</tr>
<tr>
<td>Education Introductory</td>
<td>34.9</td>
</tr>
<tr>
<td>Math Advanced</td>
<td>33.9</td>
</tr>
<tr>
<td>Psychology Advanced</td>
<td>31.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. SF-CL with Instruction</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education Advanced</td>
<td>43.0</td>
</tr>
<tr>
<td>Zoology Advanced</td>
<td>42.5</td>
</tr>
<tr>
<td>Psychology Introductory</td>
<td>41.7</td>
</tr>
<tr>
<td>Zoology Introductory</td>
<td>41.7</td>
</tr>
<tr>
<td>Math Introductory</td>
<td>41.3</td>
</tr>
<tr>
<td>Math Advanced</td>
<td>36.8</td>
</tr>
<tr>
<td>Education Introductory</td>
<td>36.7</td>
</tr>
<tr>
<td>Psychology Advanced</td>
<td>36.3</td>
</tr>
</tbody>
</table>
### D. SF-CL with Examinations

<table>
<thead>
<tr>
<th>Subject</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychology Introductory</td>
<td>41.4</td>
</tr>
<tr>
<td>Math Introductory</td>
<td>40.7</td>
</tr>
<tr>
<td>Zoology Advanced</td>
<td>40.3</td>
</tr>
<tr>
<td>Zoology Introductory</td>
<td>39.9</td>
</tr>
<tr>
<td>Education Advanced</td>
<td>37.9</td>
</tr>
<tr>
<td>Math Advanced</td>
<td>37.5</td>
</tr>
<tr>
<td>Psychology Advanced</td>
<td>35.0</td>
</tr>
<tr>
<td>Education Introductory</td>
<td>33.5</td>
</tr>
</tbody>
</table>

### E. SF-CL with General Evaluation

<table>
<thead>
<tr>
<th>Subject</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychology Introductory</td>
<td>43.0</td>
</tr>
<tr>
<td>Education Advanced</td>
<td>42.0</td>
</tr>
<tr>
<td>Zoology Introductory</td>
<td>41.8</td>
</tr>
<tr>
<td>Zoology Advanced</td>
<td>41.2</td>
</tr>
<tr>
<td>Math Introductory</td>
<td>40.6</td>
</tr>
<tr>
<td>Math Advanced</td>
<td>38.8</td>
</tr>
<tr>
<td>Psychology Advanced</td>
<td>37.9</td>
</tr>
<tr>
<td>Education Introductory</td>
<td>37.6</td>
</tr>
</tbody>
</table>
in zoology courses. The math students emulate subjects in psychology courses to a degree, i.e., introductory subjects rate the typical course higher than advanced math subjects but the extreme spread between the two ratings is not present for math subjects (44.2 for introductory and 39.1 for advanced math subjects). One thus concludes from these data that the specific Subject Field the subject is enrolled in and the Course Level of that course interact in a special fashion to produce the evaluation of the typical course.

The data were not significant for the rating of their specific courses on the Difficulty measure which effectively prevents comparisons between these two rating conditions for the same dependent variable. One might expect subjects in the introductory course to rate the typical course higher in terms of difficulty and indeed this seems to be the pattern. This tendency on the part of subjects enrolled in introductory courses is not confirmed by evaluations on the other CAF measures.

Subject Field and Course Level with Content: Typical Course. Subjects enrolled in the advanced courses who rated their courses higher on Content rated the typical course lower on the same measure (Table 14-B). This would indicate that subjects who perceive their courses stressing the higher, more abstract intellectual aspects of the subject field view the typical course as one that emphasized memorization and other mundane learning chores. Unlike the results on the Difficulty measure, Introductory subjects fail to dominate the higher levels when rating the content of the typical course. Introductory psychology and math subjects rated the typical course highest.
However subjects in the advanced zoology and education courses evaluated the same measure at only slightly lower levels.

These results suggest that subjects in the introductory level courses who rate their courses lower in terms of Content have definite expectations about other courses, i.e., those beyond the introductory level. If, for example, subjects view the present course as one where rote memory is a major emphasis they will tend to perceive other courses as stressing the higher intellectual processes such as thinking and reasoning.

One might infer from these data that the higher the subject rates the specific course on the Content measure the greater the tendency will be to view the typical course lower on the same dimension.

Subject Field and Course Level with Instruction: Typical Course. There was a tendency for subjects in the introductory courses to rate their courses lower on the Instruction measure and a concomitant trend to rate the typical course higher on the same dimension (Table 14-C). Introductory math, education, and psychology subjects had a conservative view regarding the quality of instruction in their courses. These same subjects evaluate the typical course in much more positive terms, suggesting perhaps a pattern of exceptions which enables one to endure a current monotonous or unpleasant experience via the anticipation of relief in similar situations in the future, i.e., this may be perceived as a boring class but once the fundamentals are acquired future courses will be more stimulating, quality of instruction will be higher, and so forth.
Subjects in the advanced education courses rate the typical course higher on Instruction than do subjects in other Subject Fields. The same pattern was true with their evaluation of the specific course in which they were enrolled. Superficially it would appear that these subjects appreciate the quality of instruction in their education courses and assume that their courses are representative of other courses outside the education field that they will be taking in the future. A similar pattern is apparent with subjects enrolled in introductory level education courses. However, these subjects rate their course and the typical course low on quality of instruction thereby providing support for the thesis that expectations generated from the lower level courses may generalize to the typical course. This finding is not present in the other Subject Fields on the Instruction measure. Once again these findings provide evidence of the importance of the Subject Field--Course Level interaction to understanding the scores on the outcome variables.

Subject Field and Course Level with Examinations: Typical Course.
The results of this interaction do not yield the consistent patterns discussed previously for evaluations of the typical course. The range of scores is narrower (41.4 - 33.5) and suggests a general tendency to rate the typical courses lower on this measure (Table 14-D). Subjects in advanced psychology courses and introductory education classes rate the typical course lowest on Examinations as they did when evaluating their specific courses. Introductory math and psychology subjects rated the typical course highest on Examinations, a finding quite different from the manner in which they rated their
specific courses. There is little difference in the way the advanced and introductory zoology subjects evaluated examinations, both rating the typical course higher than the mean for this data. The ratings by these two groups differ in rank from their evaluation of the specific courses in which they were enrolled. Introductory zoology subjects rated their courses very high on Examinations while those subjects in the advanced section ranked their course lower than they did the typical course. The ratings on Examinations by subjects in advance math and education courses fall in the lower half of the rankings for this measure, a position lower than their ratings of the specific courses. These results again fit the pattern discussed previously, i.e., subjects in advanced courses who rate their courses high on these dimensions tend to view the typical course in a less favorable light in comparison to their specific course.

Subject Field and Course Level with General Evaluation: Typical Course. Results from the interaction of SF-CL with the measure of General Evaluation are consistent in their diversity (Table 14-E). The findings are not consistent for either CL or SF but again are a function of the effect the combination of these two variables have on the General Evaluation measure. Psychology and education course subjects rated the typical course highest. Advanced psychology and introductory education rated it lowest. There was minimal difference in the way the introductory and zoology subjects rated the typical course nor in the ratings by the two levels of math subjects. Neither SF or CL alone explains the findings: it seems to depend upon the complex interaction between the two variables.
Interaction Effects: Difference Scores

It was hypothesized that for each main effect variable there would be differences in the way subjects would evaluate their courses, the typical course, and differences in the scores on the same CAF measures between the specific course rated and the typical course. The ensuing paragraphs are addressed to the results of the interaction effects on the difference scores. The interaction effect that is statistically significant for the difference scores is SF with CL for each of the five outcome variables on the CAF: Difficulty, Content, Instruction, Examinations, and General Evaluation. These data are presented in Table 15.

Subject Field and Course Level with Difficulty. The data from Table 16 presents an important finding from this study: most subjects rated the typical course as more difficult than their specific courses. The exceptions were subjects in the advanced zoology and advanced psychology courses who rated their particular courses higher than the typical course on the Difficulty measure. Subjects in the introductory education and introductory psychology courses provided the most extreme differences on this interaction implying that the typical course is far more difficult than the course in which they were enrolled. Other findings from this data indicate important differences in the way subjects in zoology courses perceive both their courses and the typical course. Those in advanced zoology clearly rate their course more difficult than the typical course; their peers in the introductory sections respond in the opposite direction viewing their courses as less difficult than the typical course.
Table 15
Probability Values of Analysis of Covariance F-Scores
For Interaction Effect of Predictor Variables
For CAF Measures and Difference Scores
Between This and Typical Course Ratings.

<table>
<thead>
<tr>
<th>CAF</th>
<th>SF-CL</th>
<th>SF-Sex</th>
<th>CL-Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty</td>
<td>.030</td>
<td>.868</td>
<td>.177</td>
</tr>
<tr>
<td>Content</td>
<td>.003</td>
<td>.289</td>
<td>.678</td>
</tr>
<tr>
<td>Instruction</td>
<td>.000</td>
<td>.864</td>
<td>.711</td>
</tr>
<tr>
<td>Examinations</td>
<td>.000</td>
<td>.997</td>
<td>.817</td>
</tr>
<tr>
<td>General Evaluation</td>
<td>.014</td>
<td>.985</td>
<td>.579</td>
</tr>
</tbody>
</table>

Note: Underlined scores are significant at the 105 level or greater.
Table 16

Rank Order of Mean Scores for Difference Between This Course Rating and Typical Course Rating on CAF Measures with Interaction of Subject Field and Course Level

<table>
<thead>
<tr>
<th>A. SF-CL with Difficulty</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychology Advanced</td>
<td>7.6</td>
</tr>
<tr>
<td>Zoology Advanced</td>
<td>4.3</td>
</tr>
<tr>
<td>Zoology Introductory</td>
<td>-1.2</td>
</tr>
<tr>
<td>Math Advanced</td>
<td>-1.4</td>
</tr>
<tr>
<td>Education Advanced</td>
<td>-3.6</td>
</tr>
<tr>
<td>Math Introductory</td>
<td>-7.5</td>
</tr>
<tr>
<td>Psychology Introductory</td>
<td>-12.1</td>
</tr>
<tr>
<td>Education Introductory</td>
<td>-15.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. SF-CL with Content</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math Advanced</td>
<td>20.6</td>
</tr>
<tr>
<td>Psychology Advanced</td>
<td>8.5</td>
</tr>
<tr>
<td>Education Advanced</td>
<td>7.0</td>
</tr>
<tr>
<td>Education Introductory</td>
<td>5.8</td>
</tr>
<tr>
<td>Zoology Advanced</td>
<td>0.4</td>
</tr>
<tr>
<td>Zoology Introductory</td>
<td>-1.5</td>
</tr>
<tr>
<td>Psychology Introductory</td>
<td>-2.5</td>
</tr>
<tr>
<td>Math Introductory</td>
<td>-2.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. SF-CL with Instruction</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoology Introductory</td>
<td>12.6</td>
</tr>
<tr>
<td>Education Advanced</td>
<td>9.7</td>
</tr>
<tr>
<td>Psychology Advanced</td>
<td>9.2</td>
</tr>
<tr>
<td>Math Advanced</td>
<td>8.1</td>
</tr>
<tr>
<td>Education Introductory</td>
<td>0.8</td>
</tr>
<tr>
<td>Zoology Advanced</td>
<td>0.0</td>
</tr>
<tr>
<td>Psychology Introductory</td>
<td>-4.4</td>
</tr>
<tr>
<td>Math Introductory</td>
<td>-9.7</td>
</tr>
</tbody>
</table>
### D. SF-CL with Examinations

<table>
<thead>
<tr>
<th>Course</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math Advanced</td>
<td>20.2</td>
</tr>
<tr>
<td>Zoology Introductory</td>
<td>10.5</td>
</tr>
<tr>
<td>Education Advanced</td>
<td>9.9</td>
</tr>
<tr>
<td>Psychology Introductory</td>
<td>4.0</td>
</tr>
<tr>
<td>Zoology Advanced</td>
<td>1.6</td>
</tr>
<tr>
<td>Education Introductory</td>
<td>-2.1</td>
</tr>
<tr>
<td>Psychology Advanced</td>
<td>-3.1</td>
</tr>
<tr>
<td>Math Introductory</td>
<td>-3.8</td>
</tr>
</tbody>
</table>

### E. SF-CL with General Evaluation

<table>
<thead>
<tr>
<th>Course</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychology Advanced</td>
<td>11.5</td>
</tr>
<tr>
<td>Education Advanced</td>
<td>11.2</td>
</tr>
<tr>
<td>Zoology Introductory</td>
<td>7.9</td>
</tr>
<tr>
<td>Psychology Introductory</td>
<td>6.4</td>
</tr>
<tr>
<td>Zoology Advanced</td>
<td>3.0</td>
</tr>
<tr>
<td>Math Advanced</td>
<td>1.9</td>
</tr>
<tr>
<td>Education Introductory</td>
<td>-0.8</td>
</tr>
<tr>
<td>Math Introductory</td>
<td>-7.0</td>
</tr>
</tbody>
</table>

Note: The negative (-) sign preceding scores indicates the direction of the difference is in favor of the typical course.
The absolute value of the difference scores between subjects in advanced psychology courses and those in the lower level education course is 22.6, between the course levels for psychology, 19.7 points. These differences illustrate the complexities of the evaluation process, the inappropriateness of attempting to predict outcomes, i.e., scores on evaluation measures, from major variables alone.

An important question from the above data concerns the feasibility of generalizations. Is it legitimate to conclude from these data that advanced psychology courses are the most difficult and introductory education courses the easiest or least challenging? Or can one state somewhat more definitively that from among the population sampled advanced psychology is the most difficult, advanced zoology courses next, and continuing through introductory education courses as the easiest course? These statements are not justified from the data. Each subject responds from a restricted frame of reference and thus one is limited by the data to quite specific generalizations.

In a sense this study may be affected by differences in populations which were discussed earlier in this chapter. For example the self-selection process might operate to discourage some students from entering certain subject fields because of perceived ability deficits and consequent apprehension. This would be followed by the movement of these students into fields and courses where the expectations are consistent with the students' perception of himself academically, e.g., into education. The findings from the analysis of the above data indicate specifically how subjects in courses at a given course level and subject field evaluate that course in terms of Difficulty.
and how those same subjects rate the typical course on the same measure. Subjects in advanced psychology courses rated their courses as difficult in comparison to the typical course to a greater degree than did subjects in the other courses. At the far end of the continuum subjects in the advanced psychology courses rated their courses easiest.

**Subject Field and Course Level with Content.** The data in Table 16-B point to an important result of this research. Subjects in advanced courses tend to view their courses as stressing the higher level cognitive processes such as thinking and reasoning; subjects in the introductory sections rate the typical course higher than their courses in stressing the same higher level intellectual processes.

As with the data for difference scores with Difficulty the same SF had the highest and lowest ratings. Advanced math subjects perceived a greater difference between the Content of their courses and of the typical course than other SF-CL combinations; introductory math subjects saw the difference favoring the typical course more than other subjects. Attention is directed to the point spread (23.2) between the two extreme scores and that both scores are from the same SF. It seems unusual that the extreme rating differences would both come from the same SF, in this instance math. In contrast both course levels from math and education rated the differences in a similar fashion: advanced and introductory education subjects rated their courses superior to the typical course on Content, subjects at both course levels of zoology saw little difference between the content of their courses and of the typical course.
Several implications are evident from these data on SF-CL difference scores. In general, subjects view the content of their courses as stressing thinking and reasoning to a greater extent than they believe the typical course does. This seems especially characteristic of subjects in upper level courses. Since these higher level intellectual abilities are valued in the academic environment this result speaks well for the Subject Fields included in this study. Subjects enrolled in introductory level courses, as a rule tend to view the typical course as the one emphasizing thinking and feeling while the courses in which they are enrolled stress rote memorization. This pattern is encouraging because it suggest an expectation of acquiring skills or establishing a foundation in the lower level courses and then progressing to more abstract functions and independent learning utilizing these basic skills in more advanced courses. Certainly introductory courses require a major commitment to memorization if the student is to successfully pass the course. Hence the lower rating on Content for introductory courses is not disappointing. Finally, there is a minimal difference between the scores for both zoology levels and for both education levels. One must conclude that both the introductory and advanced zoology courses demand significant effort from the student to the objective of memorizing structures and functions and therefore little apparent difference is perceived by subjects in these different course levels of the same SF. The explanation for the similarity between responses for different CLs of education SF follows a familiar tack. In part these subjects seem to acknowledge that this SF does not provide the kind
of content that requires extraordinary memorization efforts at the introductory level nor at the advanced level. These subjects seem to rate their courses as both different and better than the typical course for these reasons.

**Subject Field and Course Level with Instruction.** The results from the difference scores on the Instruction measure indicate that most subjects, regardless of SF or CL, rate their courses higher in terms of quality of instruction than they rate the typical course on the same measure (see Table 16-C). More precisely, the difference score in both direction and magnitude for subjects enrolled in introductory level zoology courses favors the course in which the subjects are enrolled over the typical course by 12.6 points (Table 16-C). This indicates that subjects in the introductory zoology courses see a difference in the quality of instruction which favors their courses to a greater extent than do subjects in other SF-CL categories on the same measure. To a lesser degree the same result characterizes, in descending order, subjects in advanced education, advanced psychology, and advanced math courses. These subjects evaluate their courses in a very positive direction on the Instruction measure which suggests that the quality of the lectures, the materials used by the teacher, and the general instructional format are viewed as significantly better than the instructional quality found in the typical course.

Subjects in the introductory zoology courses are again at the dividing point where difference score size and direction shifts to favor the typical course. These subjects and those from introductory education see little difference in the quality of instruction between
their courses and the typical course. This does not say anything about the level of that quality. Indeed these subjects may feel that both course situations are characterized by a relatively high quality of instruction. This finding simply indicates little perceived difference between the two.

The differences shift to favor the typical course when ratings by subjects in the introductory math and psychology courses are considered. These subjects clearly view the typical course superior to their own on the Instruction dimension. Perhaps to a degree not demonstrated on the other measures of the CAF the evaluation of instruction includes scales allowing the subject to evaluate the teacher as well as the other aspects of instruction. These two scores suggest the ratings may be affected by relatively poor or uninspired teaching skills.

There is pervasive order effect as a function of course level that is similar to the pattern observed with the Content measure. The subjects enrolled in advanced courses tend to rate their courses higher than the typical course while those in introductory courses tend to reverse this pattern. The one notable exception is the score from subjects in the introductory zoology courses who apparently see extraordinary differences in the quality of instruction found in their courses when compared with the typical course.

Subject Field and Course Level with Examinations. The subtleties of the SF-CL interaction and the relative factorial purity of the CAF measures are again demonstrated with these data. The direct influence of CL is less evident than with the previous SF-CL interactions with
Difficulty, Content, and Instruction measures. It is the specific combination of SF with CL that determines the outcome score in both size and direction.

Subjects in the advanced math course rate the two course situations (this and typical) in directions that portray their perceptions of examinations in their course. They rate the examinations as outstanding and far superior to those characteristic of the typical course (Table 16-D): at the opposite end of these scores are the evaluations from subjects in introductory math courses. The difference score of -3.8 indicates a discrepancy in the ratings of Examinations between introductory math courses and the typical course in favor of the latter. The subject in a lower level math course rates the typical course higher in terms of his perception of examinations. The question as to the basis for this pattern with advanced and introductory math courses is relevant. The consistently low scores for introductory math on the five CAF measures would suggest an unusual combination of subject matter, teaching style, and perhaps instructor personality that resulted in uniformly low evaluations for these courses. It is not conceivable to suggest from this data that all introductory math courses would be rated in a similar negative fashion.

The importance of the interaction is evident in the ratings by subjects in psychology courses. Introductory psychology subjects perceive their courses having better examinations than the typical course while subjects from advanced psychology saw examinations as a weakness of their courses by rating the typical course higher on this dimension.
Education courses were rated in a different direction. Subjects in advanced education courses saw a significant difference between their courses and the typical course with the former rated as much better. Subjects in the introductory education courses rated examinations in the typical courses superior to their own, a finding that suggests the subjects in lower level education courses may be closer to objective reality in rating their courses and the typical course than subjects in the advanced courses. Undoubtedly more non-education majors are enrolled in the introductory course, individuals with no emotional investment in education as a career and whose perceptions may be more accurate. It was noted earlier in this study that a sharp decline in the mean SAT-V score occurs between subjects in the introductory and those enrolled in advanced education courses. Admittedly speculative, nevertheless one might conclude that the decline in SAT-V scores was attributable to the results of their evaluation along the dimensions included in this study. These subjects may perceive education courses failing to offer the intellectual challenge, demanding examinations, and other characteristics they expect to be present in college courses, thus leaving the advanced education courses populated by students either viewing it as a means to an important objective (teaching) or those struggling to maintain their academic self-esteem by reducing the dissonance associated with enrollment in a low prestige subject field.

The difference scores for zoology courses indicate that subjects at both Course Levels evaluate their course positively in comparison to the typical course on the Examinations measure. The score for the
introductory subjects (10.5) is an indication of the quality of examinations in their courses and the relative low regard these subjects have for examinations in the typical course. Clearly these subjects believe their examinations are difficult and challenging and thus quite different from those in the typical course.

The subjects in lower level psychology courses rate their courses higher on the Examination measure than they would the typical course. The position is reversed for subjects in the advanced courses.

Subject Field and Course Level with General Evaluation. The difference scores for this interaction and the General Evaluation measure (Table 16-E) are an indication of the disparity between the subjects' evaluation of their particular courses and the perception of the typical course. The data suggest that subjects see a greater difference between the value of the advanced education and psychology courses and the typical course in favor of their own courses than other subjects making the same comparison. These students in the advanced education and psychology courses rate their courses in a positive direction and different from the typical course on this measure. In essence their courses are seen as valuable and worthwhile in a general sense in contrast to their impressions of the typical course.

In contrast subjects in introductory math courses perceive an important disparity between their courses and the typical course which indicates a general absence of positive feelings about their courses. The substance of their perceptions is in favor of the typical course as the valuable educational experience rather than the
introductory math courses.

Inspection of the data points out the order of the perceived difference between the various SF-CL interactions and the typical course. These data allow certain conclusions about subject perceptions of their courses in relation to the typical course. Most subjects rate their courses higher than the typical course. Advanced math and introductory education subjects follow this pattern to a lesser extent than other subjects. The smaller the difference scores the greater the likelihood that the general evaluation of that particular course is not especially noteworthy, e.g., advanced math (1.9), introductory education (-0.8), and introductory math (-7.0).

Discussion of Difference Scores for SF-CL Interactions. Inspection of the five significant SF-CL interactions and their difference scores on the respective CAF measures provide a basis for some tentative conclusions about the perception of subjects in the four subject fields and about the CAF as an evaluation tool.

In general these subjects have positive feelings about the courses they are evaluating. With the exception of two SF-CL combinations every course obtained a positive rating in contrast to the ratings for the typical course. This finding suggests a general level of satisfaction with the courses in which they are enrolled regardless of the SF or CL.

Students respond to the evaluation task in an objective manner. Not one of the courses was rated in a vague, general way as excellent or mediocre on all five measures. Two examples are cited to illustrate this point. The advanced psychology courses tended to obtain the highest scores on the outcome measures, i.e., were rated in a significantly
different direction from the typical course, yet on the Content and Instruction measures other courses obtained higher difference scores while the ratings on Examinations produced a negative score for the advanced psychology courses in comparison to the typical course. With each measure the discrepancy between the introductory math courses and the typical course produced negative scores pointing out the poor quality of these math courses in contrast to the typical course. On only one measure (Difficulty) were the difference scores greater in a negative direction for other courses. One might infer from such data that these particular introductory math courses possess few, if any, redeeming qualities, that in every important dimension there was an absence of the characteristics deemed important in college courses and the instruction thereof. To make such an inference is to go beyond what is warranted by the data. This result may be a function of another unexplored variable: student attitudes toward math. Advanced psychology courses generally obtained good marks from the subjects in contrast to the typical course but subjects were quick to acknowledge that examinations in those courses were not an outstanding characteristic, i.e., despite the high scores on most outcome measures these courses obtained low scores on the quality of examinations indicating that the subjects rated the typical course significantly better on this measure.

These findings emphasize the seriousness of the subjects when evaluating their courses, the relative factorial purity of the CAF, and the significant effect the combination of subject field and course level have on the outcome variables of course difficulty, content,
instruction, examinations, and the general evaluation

Other Statistical Results

The hypothesized differences in the evaluation of courses attributable to the continuous variables of expected grade in the course (GE), grade received (GR), grade Point average (GPA), and SAT-verbal and SAT-math scores were generally not supported by the data. These data are presented in Table 4 and Table 17. There was, however, support for the impact of grade expectations on the Difficulty and General Evaluation measures (Hypothesis IV), for the SAT-verbal scores and grade received on the General Evaluation measure (Hypothesis V and VI). However, the significant covariate effect was not with the evaluation of the present courses but rather with the evaluation of the typical course. This finding will be discussed in greater detail below. A discussion of the results for each covariate hypothesis follows.

Hypothesis IV

It was hypothesized that the subjects' grade expectations in the course would make a difference in the way the course was evaluated, i.e., the higher the GE the higher the ratings on the CAF measures and the greater the difference between the ratings of the present course and the typical course.

The hypothesis was confirmed for the Difficulty and Examinations measures with F-scores significant at the .0009 and .001 levels respectively. Subjects expecting to receive high grades in their courses tend to perceive these courses as more difficult, i.e., the higher the GE the more difficult the course is rated. The same pattern is found for the Examinations measure.
Although the Difficulty measure is the most independent of the five CAF factors, there would appear to be significant overlap on this hypothesis between the Difficulty and Examinations measures. It would seem reasonable to assume that the two scores found to be significant with this continuous variable are closely related. The student who expects a high grade also expects to earn that grade, i.e., to attend class, study, prepare diligently for examinations, and so forth. In essence the student's perceptions of himself in the academic context must be validated in some tangible form, preferably by a satisfactory grade in the course. The data for this study were obtained near the end of the academic term (quarter) and hence most subjects had a reasonable idea of what grade they would receive. The data thus reflect the feeling subjects had about the course in terms of the time and energy demands as well as their reactions to the examinations as a function of their performance on those devices for the assessment of learning.

A question should be addressed to the reasons for the absence of significant findings with the other outcome measures, i.e., why were significant results lacking for Content, Instruction, and General Evaluation measures. Apparently, grade expectations are related more directly to specific dimensions of a course; general difficulty level in terms of the time and effort required to master the course subject matter, and the perceived difficulty of the examinations. Note that "difficulty" is one of the defining words for both the measure of Difficulty and of Examinations. The three CAF measures without significant findings with GE involve broad and
more subjective definitions. One might conclude that the significant findings for the GE covariate with the Difficulty and Examinations measures is a reflection of the close relationship subjects perceive between high grades, difficult examinations, and the intellectual demands of the course.

**Hypothesis V**

This hypothesis asserts that the subjects' cumulative grade point average (GPA) and SAT-V and SAT-M scores would make a difference in the way the course was rated on the CAF measures. This aspect of the research was concerned with traditional academic predictors. The results for this hypothesis were limited. The higher the subjects SAT-V scores the better they rated the course on the General Evaluation measure. The F-scores are significant at the .007 level and are presented in Table 17.

The data for the SAT-verbal scores and the General Evaluation dimension are consistent with findings cited in the literature review section of this study. The absence of results with the other CAF measures is assumed to be a function of the CAF as an objective evaluation instrument. The scales upon which the five CAF measures are based effectively eliminate from the rating process any significant effect from traditional academic prediction variables such as the SAT-verbal examination. This finding is further evidence in support of a pattern observed throughout the analysis of data in this study: the relative independence and factorial purity of the CAF and the careful discriminations students make when evaluating their courses.
Table 17

Probability Values of Analysis of Covariance F-Scores For Academic Predictors with CAF Measures for This course, Typical Course, and Difference Scores Between This and Typical Course Ratings.

<table>
<thead>
<tr>
<th>CAF</th>
<th>GE</th>
<th>GR</th>
<th>GPA</th>
<th>SAT-V</th>
<th>SAT-M</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>This Course</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty</td>
<td>.000</td>
<td>.958</td>
<td>.758</td>
<td>.569</td>
<td>.195</td>
</tr>
<tr>
<td>Content</td>
<td>.510</td>
<td>.319</td>
<td>.779</td>
<td>.109</td>
<td>.378</td>
</tr>
<tr>
<td>Instruction</td>
<td>.070</td>
<td>.398</td>
<td>.346</td>
<td>.574</td>
<td>.235</td>
</tr>
<tr>
<td>Examinations</td>
<td>.001</td>
<td>.495</td>
<td>.339</td>
<td>.837</td>
<td>.157</td>
</tr>
<tr>
<td>Gen. Evaluation</td>
<td>.060</td>
<td>.013</td>
<td>.339</td>
<td>.007</td>
<td>.848</td>
</tr>
<tr>
<td><strong>Typical Course</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty</td>
<td>.819</td>
<td>.496</td>
<td>.910</td>
<td>.378</td>
<td>.398</td>
</tr>
<tr>
<td>Content</td>
<td>.957</td>
<td>.826</td>
<td>.004</td>
<td>.747</td>
<td>.476</td>
</tr>
<tr>
<td>Instruction</td>
<td>.651</td>
<td>.563</td>
<td>.000</td>
<td>.514</td>
<td>.688</td>
</tr>
<tr>
<td>Examinations</td>
<td>.993</td>
<td>.890</td>
<td>.001</td>
<td>.436</td>
<td>.979</td>
</tr>
<tr>
<td>Gen. Evaluation</td>
<td>.692</td>
<td>.590</td>
<td>.000</td>
<td>.857</td>
<td>.935</td>
</tr>
<tr>
<td><strong>Difference Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty</td>
<td>.001</td>
<td>.642</td>
<td>.746</td>
<td>.941</td>
<td>.581</td>
</tr>
<tr>
<td>Content</td>
<td>.578</td>
<td>.544</td>
<td>.039</td>
<td>.133</td>
<td>.896</td>
</tr>
<tr>
<td>Instruction</td>
<td>.067</td>
<td>.805</td>
<td>.000</td>
<td>.967</td>
<td>.508</td>
</tr>
<tr>
<td>Examinations</td>
<td>.004</td>
<td>.622</td>
<td>.000</td>
<td>.646</td>
<td>.201</td>
</tr>
<tr>
<td>Gen. Evaluation</td>
<td>.073</td>
<td>.115</td>
<td>.000</td>
<td>.023</td>
<td>.928</td>
</tr>
</tbody>
</table>

Note: Underlined scores are significant at the 105 level or greater.

GE=Grade Expected; GR=Grade Received; GPA=Grade Point Average.
Hypothesis VI

This hypothesis states that the grade the subjects receive in the course will make a difference in the way the course is rated on the evaluation measures. There was only one significant finding in support of this hypothesis: F-scores significant at the .013 level for GR with General Evaluation. This statistic is found in Table 1. This lone finding for the GR covariate requires little comment. Apparently the higher the grade received in a specific course the higher that subject tends to rate the course on the General Evaluation measure. Presumably there is a congruence between the grade received in a course and the student's perception of that course. Given a representative sample of students and courses in order to minimize the impact of the occasional "gut" course, this result seems logical. The response to queries as to the absence of significant results on the other outcome measures is again attributable to the structure and quality of the CAF.

Discussion of GPA Covariate and CAF Scores. One of the surprising and important findings of this study is the absence of any relationship between GPA and scores on the evaluation measures for the course enrolled in and the finding of significant results for GPA with the evaluation of the typical course (Table 4 and Table 17). In this study GPA did not affect evaluation of the specific courses. It had a marked impact on the subjects' perception of the typical course. F-scores were significant at .004 level with Content, .001 for Examinations, and .0009 with both Instruction and General Evaluation. The impact on the Difficulty measure was not at a significant level.
The notion that prior academic achievement history would not affect evaluation of the present course but have a definite impact on ratings of the typical course seems somewhat inconsistent. One might expect a difference between evaluations for the course enrolled in and the typical course but at least a degree of relatedness with both evaluation settings. The reason for this discrepancy is attributable to the subjects and to the instrument.

Apparently students, when evaluating a specific course, complete the process in a dedicated, purposeful manner focusing exclusively on the demands of the evaluation instrument. Their GPA at the time, other academic experiences, anything not directly related to the course being evaluated seems to be excluded. Their behavior becomes quite stimulus-specific. The stimuli are the questions on the CAF rating scales. These scales, through a long process of research, evaluation, and re-writing have been developed to such an extent that they facilitate, demand a degree of concentration and objectivity specific to the evaluation task. Thus it would seem that variables such as the GPA fail to have a direct impact on the evaluation scores.

The evaluation of the typical course is a different issue, psychologically, and therefore makes different demands upon the subject. The student is asked to perform a rather nebulous task, that of evaluating the typical course. A broader range of the student's behavioral history and expectations influences the process in this altered context. His perception of himself academically would in part be determined by his GPA at the time.
The student with the higher GPA presents a history of academic success and expectations that this pattern will continue. In order to maintain this hypothesized academic self-concept the student is required to maintain an image of the academic expectations and demands of the university consistent with his self-concept. The outstanding student in terms of GPA, will tend to perceive the typical course as one that requires ability and effort on the part of the student in order to succeed, i.e., obtain the high grade. He will thus evaluate the typical course at higher levels on the CAF measures. For example Content will be perceived as stressing the more abstract intellectual abilities of the student rather than plodding rote memorization. Examinations will be rated as challenging, difficult, and fair while Instruction would be evaluated at a similar high level. Finally, on the General Evaluation measure this student will describe the typical course as worthwhile and valuable, to be recommended to others.

Presumably the above student would rate the typical course low on the Difficulty measure. Difficulty in a sense is a relative matter. It is a function of ability, background, interest and motivation. Courses are not necessarily difficult for the high GPA student. With the above factors in mind courses that are legitimately difficult for many competent students are not perceived thusly by the high GPA student. This covariate (GPA) is a measure of ability only and hence this study cannot answer questions about the impact of non-ability variables on the Difficulty measure.
CHAPTER IV

SUMMARY

Summary

This research examined the proposition that subject field, course level, and sex of the student would make a difference in the evaluation of courses in which the student is enrolled. In addition it was suggested that traditional academic predictors such as grades, GPA, and SAT scores would affect student evaluations. The basic hypotheses were that subject field, course level, sex, grade expected, grade received, and GPA and SAT scores would make a difference in the evaluation of courses. Each hypothesis examined the ratings for the specific course the student evaluated, the evaluation of the typical course, and difference of the scores between the two ratings. Scores were obtained for five aspects of courses as measured by the Course Assessment Form, The scores were for Difficulty, Content, Instruction, Examinations, and General Evaluation. Data regarding GPA, SAT-V, and SAT-M scores, grade expectations, and grade received were obtained from the Demographic and Attitudinal Questionnaire and other institutional records. A total of 342 students participated in the study from courses at the introductory and advanced levels in psychology, math, zoology, and education courses. There were 185 female and 157 male subjects. The sample included 194 students in the introductory courses and 148 enrolled in advanced courses. The organization of students by subject fields included 115
in psychology courses, 70 in math, 97 in zoology sections, and 60 in education courses.

The statistical treatment of the data utilized an Analysis of Covariance design because of the presence of both continuous and discrete variables and in order to facilitate the examination of the effect of the important predictor variables on the course evaluations while controlling for the contribution from traditional academic predictors.

The hypotheses for the impact of subject field and course level on the specific course were supported while those for sex of the student were not. These three major hypotheses were not supported for the hypothesized impact on the typical course ratings. However, the data did verify the importance of course level to the difference scores between the two ratings and to a lesser extent that of subject field. The hypotheses for grade expected, grade received, and SAT-V and SAT-M scores were generally not supported. Although the hypothesized effect of the GPA on evaluations of the particular course did not materialize, the GPA of the student had a marked impact on the evaluation of the typical course and the difference scores between the two ratings.

Perhaps the most unique finding from this research is attributable to the rating of the typical course and the ensuing discrepancy score. It was found for example that course level did not make a difference with the rating of the typical course but the difference scores provided meaningful information in terms of how students perceived their own courses, data which may provide important information
for additional institutional research studies. It was found, for example, that students discriminated between their courses and the typical course in a selective fashion which was a function of subject field and course. There were differences within a given subject field as a function of course level and vice versa. Differences of a similar nature were also found between different subject fields and course levels. The ultimate goal of evaluation is the improvement of instruction. It would be reasonable for department chairmen and academic deans to be concerned with the kinds of data available from ratings of specific and typical courses.

It was concluded from the findings of this study that subject field and course level do make a difference in student evaluation of courses but by themselves do not affect student ratings of the typical course. Both subject field and course level predict differences between the ratings for the specific course and the typical course although Subject Field fails to be effective with Difficulty and Instruction scores. The interaction effect of subject field with course level is effective with virtually every outcome measure across all three conditions: specific course, typical course, and difference score. Grade expected was related to the Difficulty and Examination ratings for the course evaluated and for the difference scores for these two measures. GPA was not related to the ratings for the specific course but, with the exception of the Difficulty score, was related to the evaluation of the typical course and the difference scores. The sex of the student had no measurable affect on the ratings. None of the interactions involving the students sex was significant. With the exception of
GPA and the typical course the traditional academic predictors were not related to evaluation of specific courses, typical course, or the differences between the two ratings. It is suggested that the structure of the evaluation instrument is in part responsible for the absence of findings with these academic predictors.

The lack of significant results attributable to sex may be a function of the sample or a reflection of increasing evidence that many sex-role stereotypes and the attendant reinforcement of attitudes and values are being eroded. The developmental studies summarized by Maccoby and Jacklin (1974), the research of Lunneburg (1972), and the review of literature by O'Leary (1974) provide stimulation for further research supporting the absence of sex differences in areas such as the ways in which males and females evaluate courses.

The thrust of this research is on student evaluation of courses, their perception of that experience rather than the evaluation by teachers or administrators. It does not measure teachers' or students' personality except inferentially. The perceptions of the course experience were measured by the CAF. The control of the intrusion of personality variables into the evaluation process is a function of this evaluation instrument. The reliability data in support of the CAF is impressive; the findings from this study are further evidence in support of the validity of this instrument: It purports to measure student perceptions of the educational environment, i.e., college courses and it accomplishes this quite precisely with measures of course difficulty, content, instruction, examinations, and a general evaluation. It excludes to an important extent the influence of both traditional academic predictors and teacher personality from the rating process.
This is because the items comprising the instrument are tied to specific behaviors.

**Limitations**

Several limitations of this study may be noted. The subjects were American college students at a small, expensive liberal arts institution in the Midwest. Approximately 65 percent of the student body are from the affluent suburbs of major cities in the New England-Middle Atlantic area. A sizeable minority graduated from private high schools (prep schools). Thus, the population from which the sample for this research was drawn was more heavily white, affluent, upper middle class than one would expect at state universities. The class sections rated were generally small and in two instances preventing more extensive data analysis because of the limited number of subjects (males in advanced education and females in advanced zoology). The evaluation instrument (CAF) is designed to assess student perception of a particular experience of which he is a part—a college course. But the very precision of the CAF limits the assessment of the course experience, i.e., certain variables such as teacher personality are inferred but not directly measured. In order to obtain more data regarding this total experience other measures would need to be included.

**Future Research**

It would be important to conduct further studies investigating variables related to evaluations of courses by students. A necessary shift must be toward more experimental studies rather than continued reliance upon correlational research. The recent study of Holmes
(1972) cited in a previous chapter sets an important precedent because he focused on student expectations and what impact these expectations would have on evaluations of the course. Replication of this study utilizing students and courses at institutions with different characteristics (state universities in the South, Far West, etc.) may provide data that supports or refutes the findings of this research. A carefully designed study with equal numbers in each design category would be useful.

The final recommendation would be for a comprehensive study based on a theoretical model of evaluation. To this point theory has not been the basis for research. There have been general statements articulated referring to the importance of theory but these suggestions are specific to a broad theoretical base for an evaluation philosophy of the institution per se. It is the recommendation of the present writer that small, carefully designed studies examining a specific aspect of a theoretical network be carried out, that the research involving a major segment of the institution must be preceded by the limited studies providing a sound theoretical basis for the more extensive variety.
APPENDIX A

COURSE ASSESSMENT FORM
### COURSE ASSESSMENT FORM

**Frank M. Fletcher**  
**The Ohio State University**

<table>
<thead>
<tr>
<th>COURSE NO.</th>
<th>GRADE IN COURSE</th>
<th>ACADEMIC STANDING</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### MAKE JUDGMENT ON THE BASIS OF WHAT THIS COURSE MEANT TO YOU.

- Harner than most courses
- Routine thinking
- Critical thinking
- Good lectures
- Poor lectures
- Inadequate exams
- Well constructed exams
- Sorry I took course
- Glad I took course
- Logical reasoning
- Ordinary retention
- Time consuming
- Minimal studying
- Unimaginative instructor
- Creative instructor
- Recommend to others
- Will not recommend
- Poor coverage in exams
- Good coverage in exams
- Little effort required
- Extra effort required
- Instructor above average
- Instructor below average
- Scholarly content
- Stereotyped content
- Appropriate exams
- Little application
- Improving instruction
- Unimpressive instructor
- Stimulating instructor
- Uninspiring
- Pertinent to everyday life
- Little application
- Leisurely course
- Strenuous course
- Learning content
- Challenging content
- Meaningful course
- Course lacks purpose
- Poor exams
- Good exams
- My toughest course
- My easiest course
- Lectures poorly presented
- Lectures well presented
- Stress on understanding
- Stress on memorizing
- High quality exams
- Low quality exams
- Unimportant course
- Significant course
- Much work demanded
- Little work demanded
- Rate memory
- Reflective consideration
- Exams harder than most
- Exams below average
- Routine Instruction
- Challenging instruction
- Worthless course
- Valuable course
- Memory emphasized
- Reasoning required
- Difficult "C" course
- Easy "A" or "B" course
- Poor Instructor
- Excellent instructor
- Easier for more
- Have had enough
- Exams helpful
- Exams distracting
- Less studying than usual
- More studying than usual
- Low quality lectures
- High quality lectures
- Mastery of ideas
- Retention of facts
- Unfair exams
- Fair exams
- Clear course objectives
- Vague course objectives
- Easier than expected
- Harder than expected
- Exams emphasized thinking
- Exams emphasized memory
- Routine Instruction
- Challenging instruction
- Waste of time
- Worth taking
- Exams detracted from course

#### PLEASE PRINT NAME

**LAST**  
**FIRST**

**MIDDLE**

---

Consider each scale a 4-letter judgment.

Make one scale only for each scale. Mark every scale. Use pencil only.
APPENDIX B

DEMOGRAPHIC AND ATTITUDINAL QUESTIONNAIRE
NAME: ________________________________

Mark only one response for each question.

1. Sex: Male Female
2. Marital Status: 1-Single; 2-Engaged; 3-Married; 4-Other
3. How many children (including yourself) are there in the family? ___
4. Are you an oldest child, but not an only child? Yes No
5. Are you a youngest child, but not an only child? Yes No
6. Which of the following statements best describes your father's education?
   1-eight grades or less
   2-attended high school
   3-graduated from high school
   4-attended college
   5-received a bachelor's degree
   6-received a master's degree
   7-received a professional degree (law, medicine, etc.)
   8-received a doctorate
7. Which of the above statements best describes your mother's education? ______
8. Which of the above statements best describes your ultimate educational goal? ______
9. With which of the following American subcultures could you most easily identify yourself? 1-Protestant; 2-Catholic; 3-Jewish; 4-Other.
10. Of the two, which of the following could you most easily identify yourself with? 1-Democrat; 2-Republican
11. How many grade schools did you attend? _____
12. How many high schools did you attend? _____
13. Did you ever attend a college other than this one? Yes No
14. What is your official class rank? 1-Fr; 2-So; 3-Jr; 4-Sr; 5-Other.
15. Are you now working while attending school? Yes No
16. Have you ever been on academic probation (because of low grades)?
   Yes No
17. Have you ever been on disciplinary probation? Yes No
18. What is your major field of academic specialization? _____________
   If undecided - what field do you expect to major in or in what direction are you currently leaning? ________________
19. If you had it all to do over again, would you choose to attend this school? Yes No
20. How many children do you want to have? _____
21. In real of constant dollars, do you expect to have a family income that is:
   1-much less than your parents
   2-less than your parents
   3-about the same as your parents
   4-more than your parents
   5-much more than your parents
22. How many units of college credit have you had prior to this term? _
23. How many units of college credit have you had prior to this course in the same subject matter area? 

24. What do you expect your grade in this course to be?  
   1-A; 2-B; 3-C; 4-D; 5-F 

25. Which of the following statements best describes the reason for your enrollment in this course? 
   1-Required course in major area. 
   2-Required course for major, but outside of major area. 
   3-Fulfills a distribution requirement, but this course not specifically required. 
   4-An elective. 

Possible responses for the items below:  (items 26 through 34) 

10% 
20% 
30% 
40% 
50% 
60% 
70% 
80% 
90% 

What percent of the students in this course do you believe would be equal to or below you in:  

26. Course grade ____ 
27. Cumulative grade point average ____ 
28. Ability in English ____ 
29. Ability with Mathematical concepts ____ 
30. Ability with Social Science concepts ____ 
31. Ability with Natural Science concepts ____ 
32. General ability ____ 
33. Making full use of ability for academic achievement in general ____ 
34. Making full use of ability for academic achievement in this course ____

THANKS FOR YOUR TIME AND COOPERATION !!!

RNS370
LIST OF REFERENCES


Bendig, A. W. The relation of level of course achievement to students' instructor and course ratings in introductory psychology. *Educational and Psychological Measurement*, 1953, 13, 437-448.
Blum, M. L. An investigation of the relation existing between students' grades and their ratings of the instructor's ability to teach. *Journal of Educational Psychology*, 1936, 27, 217-221.


Fletcher, F. M. Student assessment of an educational experience. Unpublished manuscript, The Ohio State University, 1972.


Harris, M. B. Sex role stereotypes and teachers evaluations. *Journal of Educational Psychology, 1975, 67*, 751-756.


Holmes, David S. Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructor. *Journal of Educational Psychology, 1972, 63*, 130-133.


Slobin, D. Y., & Nichols, D. G. Student rating of teaching. *Improving College and University Teaching,* 1969, 17, 244-248.

Thompson, Eugene W. A study of the discrepancies between student evaluation and faculty self-perceptions of instructional procedures in higher education. Western Michigan University, 1972.


