Sensitivity of the Comp Objective Test to Structural Differences Among General Education Programs in Liberal Arts Colleges and Comprehensive Universities

Christine Ann Brooks
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.
SENSITIVITY OF THE COMP OBJECTIVE TEST TO STRUCTURAL DIFFERENCES AMONG GENERAL EDUCATION PROGRAMS IN LIBERAL ARTS COLLEGES AND COMPREHENSIVE UNIVERSITIES

by
Christine Ann Brooks

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Education
Department of Educational Leadership

Western Michigan University
Kalamazoo, Michigan
June 1994
The purpose of this study was to investigate the validity of the College Outcome Measures Program (COMP) by determining if COMP is sensitive to differences among general education programs.

Two variables provided the operationalization of general education: (1) degree of institutional control over general education course selection, and (2) percent of total coursework devoted to general education. To control for additional factors known to impact upon students' learning during college, four concomitant variables were used in the analysis. These variables were selected following a review of the literature on the effects of college. The level of analysis was the institution; forty-one institutions comprised the final study sample. Least squares linear regression was used to analyze the data; a hierarchical procedure was used to enter variables into the regression models.

Since neither of the two general education variables produced statistically significant slopes in any of the regression models, neither of the two null hypotheses that formed the framework for the analysis were rejected. Therefore, there is no certainty that COMP is sensitive to differences among programs as defined by the degree of control that an institution assumes over general education course selection or the relative number of general education courses that students complete. However, both the sample size and the effect size in this study were small, meaning...
that statistical power was very low.

Despite the fact that the null hypotheses could not be rejected, two interesting findings resulted from the analysis. First, percent of students that are transfers produced statistically significant slopes in four of the seven regression models, suggesting that the degree of transfer behavior in an institution affects educational programming for all students. Second, several fairly small but noticeable partial correlations were found between the general education characteristics and COMP subscales. These two findings suggest that further validity research needs to be done on COMP using an expanded definition of general education. An additional benefit of this research is the contribution it makes to the literature on methodologies for studying the validity of tests designed to assess programs.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Sensitivity of the COMP objective test to structural differences among general education programs in liberal arts colleges and comprehensive universities

Brooks, Christine Ann, Ed.D.
Western Michigan University, 1994
ACKNOWLEDGMENTS

Many people contributed assistance and support during the completion of this dissertation. Of those who played a major role, I must first mention my advisor and dissertation director, Dr. Mary Anne Bunda, who was the first to teach me the artful subjectivity of quantitative analysis. Her unrelenting standards of educational rigor were only matched by my unrelenting use of e-mail. For her support, consideration, and high response rate, I will be forever grateful. My committee members, Dr. Uldis Smidchens and Dr. John Crowell, taught me more than I thought I ever wanted to know about statistics. They should each know that I thoroughly enjoyed every minute.

Secondly, a sincere thanks goes to Dr. Joe Steele of the American College Testing Company who not only provided data (rather important!) and COMP documentation, but also led me to several key articles in the area of general education.

Thirdly, I cannot thank enough my friends both in Indiana and, more recently, in Maine who encouraged me in many ways to complete this dissertation.

Finally, I must acknowledge two people who gave me many special gifts and taught me many important lessons, all of which were put to use many times over during my years in graduate school. I dedicate this dissertation to the memory of my mother and father.

Christine Ann Brooks
# TABLE OF CONTENTS

ACKNOWLEDGMENTS........................................................................................................... ii

LIST OF TABLES................................................................................................................ vi

LIST OF FIGURES.............................................................................................................. viii

CHAPTER

I. STATEMENT OF THE PROBLEM......................................................................................... 1
   Assessment in Higher Education................................................................................. 1
   The College Outcome Measures Program.............................................................. 3
   Purpose of the Study............................................................................................... 7

II. REVIEW OF THE LITERATURE..................................................................................... 9
   Research Assumptions.......................................................................................... 10
      Assumption One: The Relatedness of Student Change and Educational Programs... 10
      Assumption Two: Legitimate Use of Tests to Assess Programs................................ 14
   Validity of COMP as a Measure of Program Effectiveness..................................... 17
   Development and Description of the COMP Instrument....................................... 18
   Studies of the Validity of COMP........................................................................... 21
   Conceptual Framework for Further Studies of the Validity of COMP....................... 31
   Defining the General Education Program............................................................. 33
   Additional Variables That Affect Learning............................................................ 35
   Studying the Validity of COMP............................................................................ 37

iii

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table of Contents--Continued

CHAPTER

III. STUDY DESIGN AND METHODOLOGY ............................................ 38
    The Conceptual Model ........................................................................ 38
    Design of the Study ............................................................................ 40
        Variables ...................................................................................... 41
    Level of the Data ........................................................................... 47
    Study Population and Sample ........................................................ 48
        Inherent Design Problems ............................................................. 50
    Data Collection .................................................................................. 52
    Analysis Procedures ............................................................................ 55
        Exploratory Data Analysis ............................................................ 55
        Multiple Regression Analysis ....................................................... 55
        Examination of Outlying Observations ......................................... 57

IV. RESULTS OF THE ANALYSIS ........................................................... 58
    The Final Sample ................................................................................ 58
    Exploratory Data Analysis .................................................................. 60
        Descriptive Statistics of Dependent Variables ............................... 61
        Descriptive Statistics of Independent Variables ......................... 62
        Descriptive Statistics of Concomitant Variables ............................ 65
        Zero-Order Correlations ................................................................ 66
    Multiple Regression Analysis ............................................................. 69
        Model One: Explaining Differences in COMP Total Means .......... 72
        Model Two: Explaining Differences in the Functioning Within Social Institutions Subscale Means .............................................. 75
Table of Contents--Continued

CHAPTER

Model Three: Explaining Differences in the Using Science and Technology Subscale Means..................................... 78

Model Four: Explaining Differences in the Using the Arts Subscale Means..................................................... 80

Model Five: Explaining Differences in the Communicating Subscale Means................................................... 81

Model Six: Explaining Differences in the Solving Problems Subscale Means................................................ 83

Model Seven: Explaining Differences in the Clarifying Values Subscale Means........................................... 84

Summary of Regression Results.................................................... 86

Examination of Outlying Observations............................................... 90

Summary of Results........................................................................... 92

V. CONCLUSIONS................................................................................... 93

Discussion of the Results................................................................... 94

Strengths and Weaknesses of This Study........................................... 99

Final Remarks.................................................................................... 100

APPENDICES

A. Release Letter Signed by Participating Institutions...................... 102

B. Letter From Human Subjects Institutional Review Board............. 104

C. Structure of General Education Questionnaire............................ 106

D. Sample Residual and Probability Plots........................................ 109

E. Externally Studentized Residuals.................................................. 115

F. Cook's Distance Values.................................................................. 118

BIBLIOGRAPHY............................................................................................... 121
LIST OF TABLES

1. Pre-Selected Study Variables ................................................................. 42
2. Sources of Independent and Concomitant Variable Data ....................... 54
3. Attributes of the 41 Institutions Comprising the Final Sample ................ 59
4. Descriptive Statistics for Dependent Variables for N = 41 Colleges and Universities .......................................................... 62
5. Descriptive Statistics for Independent Variables for N = 41 Colleges and Universities .......................................................... 63
6. Descriptive Statistics for Concomitant Variables ................................... 66
7. Zero-Order Correlations Between Dependent Variables and Other Variables .......................................................... 67
8. Zero-Order Correlations Among Concomitant and Independent Variables ..................................................................................... 68
9. Hierarchical Analysis Results of COMP Total Means ............................ 71
10. Regression Estimates of COMP Total Means ......................................... 73
11. Hierarchical Analysis Results of Functioning Within Social Institutions (FSI) Means .......................................................... 76
12. Regression Estimates of Functioning Within Social Institutions (FSI) Means .......................................................... 77
13. Hierarchical Analysis Results of Using Science and Technology (US) Means ..................................................................................... 78
15. Hierarchical Analysis Results of Using the Arts (UA) Means ................. 80
16. Regression Estimates of Using the Arts (UA) Means .............................. 81
17. Hierarchical Analysis Results of Communicating (COM) Means ............ 82

vi
List of Tables—Continued

18. Regression Estimates of Communicating (COM) Means ........................................ 83
19. Hierarchical Analysis Results of Solving Problems (SP) Means .............................. 84
20. Regression Estimates of Solving Problems (SP) Means ........................................ 85
21. Hierarchical Analysis Results of Clarifying Values (CV) Means ............................. 86
22. Regression Estimates of Clarifying Values (CV) Means ........................................ 87
23. Summary of Regression Analysis ........................................................................... 88
LIST OF FIGURES

1. The Six Outcome Areas Measured by COMP ............................................. 19
2. Conceptual Model for a Study of Effects of General Education Program Structure Upon COMP Scores ................................................................. 39
3. Bar Graph of Degree of Institutional Control Over General Education Course Selection .................................................................................. 64
4. Bar Graph of Percent of Total Coursework Devoted to General Education ........................................................................................ 64
CHAPTER I

STATEMENT OF THE PROBLEM

Assessment in Higher Education

One of the ironies of recent times is that higher education, while appearing to place great intrinsic value on knowledge, possesses shockingly little knowledge of itself (Ewell, 1984). The rhetoric of academia gives high regard to the seeking of knowledge for the sake of understanding and self-improvement; colleges and universities, particularly those focusing on the liberal arts, aim to instill in their students this same appreciation for knowledge and self-improvement. Yet, until recently, institutions of higher education have not made it a priority to seek knowledge about themselves in order to understand and improve the quality of teaching and learning. Given recent external demands for accountability and internal demands for reform (Erwin, 1991), many colleges and universities are now looking within to gain an understanding of their programs and to make improvements where shortcomings are found.

Over the last 10 to 15 years, there have been public calls for accountability from state governments (Barak, 1982), a growing interest in the effects of college on the part of accreditation associations (Manning, 1986), and a push for curricular reform from higher education practitioners (Association of American Colleges, 1985). In response, college and university faculties and administrators have engaged in long conversations about the objectives of the curriculum, the meaning of a college degree, and the effects, if any, of completing a college education. The
result has been a shift in the way that institutions think about quality and effectiveness. Once focused on resources, such as the number of volumes in the library, number of faculty with doctoral degrees, and the dollar value of laboratory equipment, colleges now consider the importance of the outcomes, that is, the knowledge and skills attained by students during college, the proportion of students who complete their college degree, and the eventual career success of college graduates (Ewell, 1983). Pascarella and Terenzini (1991), in their extensive review of the literature on the impact of college upon students' cognitive, psychosocial, and moral development, suggest that there are meaningful associations between student outcomes and institutional policies and programs. Thus, concerns for how students are affected by going to college are increasingly accompanied by an interest in program evaluation and policy analysis.

The classic rational model of organizational process postulates that people in organizations rely on information for assistance in choosing an appropriate course of action (Ewell & Chaffee, 1981). Consequently, colleges and universities that focus on understanding and self-improvement necessarily seek information to guide the discussions of college outcomes. Discussions about effectiveness in the "self-regarding" institutions (Ewell, 1984) rely on information about what students are experiencing and learning through the college curriculum and co-curricular programs and activities. Given such information, internal evaluation efforts can occur through which judgments about programs can be made, policies developed, and improvements put in place (Bunda, 1988). Recent interest in higher education assessment is rooted in this need for information upon which to base decisions about programs.

Whether initiated in response to a mandate from state government or an internal desire to understand and improve, assessment has forced people to think
about how best to conceptualize and measure the effects of college. After all, if something is not measured, it is less likely to be improved (Cross, 1990). Assessment practices are based on the notion that by measuring student outcomes, educators will have the information they need to make judgments about programs, develop policy, and, in turn, improve the quality of educational programs and practices. Ultimately, assessment is attempting to apply the rhetoric of academia, that is, the seeking of knowledge for the sake of understanding and self-improvement, to things at the very center of an academician's day-to-day life: the curriculum, the classroom, and the process of teaching and learning.

The College Outcome Measures Program

In 1976, the American College Testing (ACT) company initiated development of their College Outcome Measures Program (COMP) instruments for the purpose of enabling colleges and universities involved in assessment efforts to gather the evidence needed for judging the quality of their general education programs. Development of COMP was based on the premise that the traditional indices of academic achievement—college degrees, final examinations, course grades, and grade point averages—are not sufficient indicators of what students have learned and how well they have been prepared for adult life. The COMP was specifically developed to provide an appropriate means for colleges and universities to assess the knowledge and skills that undergraduate college students are expected to acquire as a result of general education programs (Forrest & Steele, 1982). To date, several hundred colleges and universities have used COMP to monitor their outcomes in order to sustain or improve delivery of general education programs (American College Testing Program, 1991a).

Issues of test validity are critically important to users of tests such as COMP.
The crux of a test's validity lies in the degree to which inferences and actions based on test scores are correct and appropriate (Messick, 1987). As a result of a college's use of COMP, information is available to guide faculty and administrators in making judgments and changes in the general education program. However, if the information is not valid, if it cannot be relied upon to lead to correct and appropriate inferences and actions, the judgments and changes may be misguided and wrong. Therefore, the validity of COMP is of great importance to users of the instruments.

Unlike tests designed to assess individuals (such as the Scholastic Aptitude Test or SAT), COMP was developed to assist institutions with the assessment of their programs, specifically general education programs. Studies done to support or refute the validity of COMP must in some way address this purpose. As such, comprehensive evidence of COMP's validity needs to include an examination of two critical areas: (1) the degree to which COMP adequately reflects the goals and content of the undergraduate general education program, and (2) the degree to which COMP accurately measures the general education knowledge and skills of groups of students who have been exposed to a particular general education program. Both of these in turn affect the degree to which COMP provides meaningful information upon which college and university faculty members and administrators can make decisions to improve programs. Traditional approaches to validity studies were developed in an era when tests were solely used to examine differences among individuals, not programs. Adequate study of the validity issues relating to COMP requires that researchers reexamine validity research methodologies; those that worked in the past may not have meaning for instruments designed to provide institutional decisionmakers with the information they need to understand and improve programs.

Over the last 15 years, researchers have attempted to study various aspects
of the validity of COMP. Most of these validity studies can be grouped within one of two types of methodologies. First, there are studies that have examined COMP's content validity, that is, the extent to which the tests reflect the goals and content of the general education curriculum of a particular college or university. These studies have relied on the method of using panels of experts to compare COMP objectives or individual items with written statements of general education goals and course content. Second, there are studies that have used correlational methodologies to look at the extent to which COMP scores are associated with various student-level characteristics such as gender, age, college-entrance test scores, or college grade point averages (GPAs). The goal of these studies has been to determine the nature of the construct measured by COMP, particularly as it might be associated with relevant individual student traits.

Several conclusions can be drawn from these validity studies. First, there is general acceptance of the theoretical assumption that COMP is testing at least some of the outcomes that are held in common by nearly all college and university general education programs (McGuire, 1987; McLean, Bolland, Rogers, Ernest, & Bullard, 1989; Pike & Banta, 1989; American College Testing Program, 1992). Second, COMP does not appear to be sensitive to gender, ethnic origin, or age differences (Forrest & Steele, 1982; Pike, 1989a). Third, COMP scores can in part be explained by the pre-college achievement levels of students as measured by the SAT or the American College Test (ACT) (Forrest & Steele, 1982; Pike, 1989a). Fourth, COMP scores correlate with length of exposure to higher education and college GPAs (Forrest & Steele, 1982; Schomberg, Hendel, & Bassett, 1982; Sibert & Ayers, 1989; Olsen, 1990). Fifth, COMP scores appear to be sensitive to the types of courses that students complete, with, for example, students scoring higher on the COMP subscale that relates directly to their academic major (Forrest &
Steele, 1982; Pike, 1989a; Pike & Banta, 1989).

While these conclusions lend some credence to the validity of COMP, they do not adequately address the second of the two critical issues in COMP validity: the degree to which COMP accurately measures the general education knowledge and skills of groups of students who have been exposed to a particular general education program. As stated earlier, there is strong evidence that there are connections between institutional programs and policies and student change (Pascarella & Terenzini, 1991). Therefore, it is reasonable to assume that different general education programs produce different results in students' learning. It then follows that the fundamental validity question posed by the issue above is whether or not COMP is sensitive to differences among general education programs. If it is, COMP can appropriately be used to measure change in students, and, by attributing this change to general education programs, COMP scores can be used to make decisions about improvements in these programs.

As part of assessment efforts, institutions are known to be using COMP to judge, improve, and monitor changes in their general education programs (Yarbrough, 1991). In addition, by providing group norms, the ACT company encourages institutions to regularly compare their scores with those of others (American College Testing Program, 1991b); by so doing, each college is inferentially comparing the quality of its program with that of others. All of this suggests that the validity of COMP must be addressed from the viewpoint of its ability to be sensitive to differences among programs. Indeed, Yarbrough (1992) considers the investigation of program effects to be of extreme importance in obtaining needed validity evidence for using or not using COMP for program evaluation.

There are four validity studies that have attempted to isolate the effects of
different general education programs or changes in programs over time. The first study found that COMP scores of students who had completed a baccalaureate degree were higher than those of vocational program completers; the former group had completed a typical four-year general education component while the latter group had experienced no general education component (Forrest & Steele, 1982). The second study utilized data from 44 different institutions and found that COMP scores seem to be related to several characteristics of general education programs, for example, the amount and breadth of general education coursework required of students at certain institutions (Forrest, 1982). The third study found some differences in COMP scores between students who had completed different baccalaureate programs within the University of Minnesota (Schomberg, Hendel, & Bassett, 1982). The fourth study, using a case study approach, found that changes in general education programs within particular institutions seemed to result in changes in COMP scores (Steele, 1989). These studies provide some preliminary evidence that COMP is sensitive to the effects of different general education programs, yet each has some shortcoming in design or implementation that makes it difficult to conclude that COMP is a valid measure of program effects. These studies will be described in greater detail in Chapter II.

Purpose of the Study

The presupposition guiding this study is that the COMP instruments are valid to the extent that COMP scores are sensitive to differences among general education programs; without this sensitivity to program differences COMP cannot be regarded as useful to institutions working to assess their general education programs. The purpose of this study is to contribute empirical evidence that either supports or refutes the validity of COMP in light of this presupposition.
The problem being addressed by this study, namely, the extent to which COMP scores are explained by differences among programs, fits squarely into the discipline of measurement science as a study of test validity. The validity of assessment instruments, however, cannot be confronted purely as a measurement problem; it, too, is an evaluation problem. Not only are individual colleges and universities using scores to make judgments about programs and improve programs, but university systems and state governments are using scores to draw conclusions about the quality of programs and, in some cases, to even award funding (Banta, 1988). There is clear evidence that COMP is being used for institutional evaluation.

Considering the known uses of the COMP instruments, the measurement issues relating to COMP quickly meld into the evaluation issues relating to COMP. An awareness of the connection between validity research and program evaluation is not new (Cronbach, 1988), yet nowhere, perhaps, has it been more apparent than in the study of instruments being used for program assessment. The central idea that must be maintained is that for COMP to be a useful assessment tool, it must be both technically sound and be able to provide the type of information that colleges and universities need to make decisions about programs. Therefore, the disciplines of measurement and evaluation must come together to provide the needed assurances that a particular instrument is both valid and beneficial as an assessment tool. Only with these assurances can informed discussions take place that actually lead to improvements in higher education.
CHAPTER II

REVIEW OF THE LITERATURE

The COMP is one of several instruments currently being used by colleges and universities to provide information that assists faculty and administrators with the evaluation of educational curricula and various aspects of student life. The COMP was specifically designed to be used to evaluate the general education program by assessing the knowledge and skills attained by students as a result of completing general education programs (Forrest & Steele, 1982). The specific concern of this study is with the validity of COMP for this recognized purpose.

The validity of COMP as a program assessment instrument rests primarily on two issues: (1) evidence of content validity; and (2) evidence of COMP's sensitivity to differences among programs (Yarbrough, 1992). The first issue concerns the fact that COMP, to be a valid program assessment instrument, must reflect the goals and content of college and university general education programs. Several research studies have been done that have contributed fairly convincing evidence of COMP's content validity (McGuire, 1987; McLean, Bolland, Rogers, Ernest & Bullard, 1989; Pike & Banta, 1989; American College Testing Program, 1992). The second issue underlying COMP's validity is that the test scores must be sensitive to differences among programs; if colleges are to rely on comparative or historical COMP data to make inferences about program quality, to make decisions about program content, and to monitor program effectiveness, COMP's validity as a measure of program effects must be determined. While a few studies (Forrest &
Steele, 1982; Forrest, 1982; Schomberg, Hendel & Bassett, 1982; Steele, 1989) have attempted to isolate the effects of program differences upon COMP scores, this second issue regarding COMP's validity as a program assessment instrument has not yet been sufficiently researched.

The general purpose of this study is to contribute to the literature on the validity of COMP by examining the test's ability to be sensitive to specific differences among general education programs. Two assumptions, supported by the educational and psychometric literature, provide the conceptual foundation for this study. First, it is assumed that student change and educational programs and policies are interrelated; institutions that wish to affect change in students properly assess programs to improve and monitor the education of their students. Second, tests such as COMP can legitimately be used to assess the impact of programs upon student change, but evidence of test validity must be provided that is based upon this use of the instrument.

The purpose of this chapter is, first, to provide support for these two assumptions and, second, to review the literature on previous validity studies of COMP. In conclusion, it will be shown that additional research is needed that examines COMP's ability to be sensitive to differences among general education programs. A conceptual model for conducting this type of research will be presented.

Research Assumptions

Assumption One: The Relatedness of Student Change and Educational Programs

This study is in part based on the assumption that differences in educational programs are associated with differences in student outcomes. Colleges and
universities that wish to have a positive impact upon student learning and development can benefit from evaluating their programs and seeking ways to improve educational practices and policies. Recent interest in higher education assessment has come about largely because of a recognized need for information upon which to base decisions about programs.

The works of Philip Jacob (1957) and Everett Eddy (1959) were significant starting points in the study of college impact. Both investigators used multi-institutional samples and examined broad outcomes of college attendance. Jacob's work focused on the effects of the curriculum, the instructor, and teaching methods upon students' values; Eddy's work looked at students' characters, that is, the ways in which students' moral principles guided their conduct as a result of being in college. These two works led a generation of future social science researchers in their examination of this critical question: in what ways and to what extent does college bring about change in students?

Since the 1950s, researchers have looked at the ways in which college attendance influences cognitive learning, attitudes, values, life and career goals, and income. Two comprehensive summaries of the outcomes literature have been published. Feldman and Newcomb (1969) reviewed and synthesized over 1,500 studies for their landmark book, The Impact of College on Students. In How College Affects Students, Pascarella and Terenzini (1991) updated this earlier work and incorporated discussion of several student development theories and models that were advanced during the 1970s and 1980s.

Questions that have been pursued in student outcomes research include: (a) how do students change as a result of going to college? (b) what are the effects of particular institutional characteristics and curricular or co-curricular programs upon student growth and development? (c) do institutions differ in their impact upon
students' learning, attitudes, social skills, and moral development?

In general, the literature on student outcomes suggests three points. First, students do change as a result of going to college. Studies that have compared young adults who have attended college with those who have not gone to college consistently report significantly greater intellectual achievement among those who have attended college (Robertshaw & Wolfle, 1983; Wolfle, 1983, 1987). Second, after controlling for students' pre-college level of achievement, student change appears to be only somewhat associated with structural/organizational characteristics of institutions such as public versus private control, budget parameters, faculty-student ratio, number of faculty with doctoral degrees, number of library books, or size of the institution (Nichols, 1964; Astin, 1968; Astin & Panos, 1969; Rock, Centra & Linn, 1970). Third, student change does, instead, appear to be convincingly associated with what colleges do programmatically. Characteristics of the curriculum, advising, residence life, and student activities appear to be the most important considerations in understanding the effects of going to college (Centra & Rock, 1971; Chickering, 1974; Pace, 1984; Pace, 1990). The critical implication is that individual colleges and universities can shape the educational and psychological outcomes of their students by promoting efforts to evaluate and improve programs (Pascarella & Terenzini, 1991).

Particularly important to this investigation into the validity of COMP are studies that have found differences in curriculum to be related to differences in the cognitive and intellectual outcomes of students:

1. Comparing data from seven colleges, Dressel and Mayhew (1954) found that the greatest first-year gains in critical thinking occurred at colleges with courses specifically organized for general education purposes and with definite core requirements involving the completion of all or a major portion of these courses in
the first year of college. These researchers also found that the amount of exposure
to science courses within specific institutions was significantly and positively
associated with first-year gains on a critical thinking measure based on science
content; additionally, the amount of exposure to humanities courses was
significantly and positively associated with gains on a critical thinking measure
based on humanities content.

2. Rock et al. (1970) found that Graduate Record Examination (GRE)
scores residualized for the effect of pre-college achievement were positively
associated with measures of the amount of faculty-student interaction, curriculum
flexibility, cultural opportunities, and high levels of challenge in academic
coursework. Centra and Rock (1971) found the same to be true for academic
achievement that was measured by separate humanities, natural science, and social
science tests.

3. Ratcliff and Associates (1988) found that senior-year measures of
learning (GRE) controlled for pre-college academic achievement (SAT) appeared to
be associated with particular patterns of coursework taken throughout the four years
of a college education. For example, a course cluster consisting primarily of upper-
division business and social science courses was associated with high achievement
on analytical reasoning and mathematics items, while a course cluster consisting
primarily of lower-division courses in the arts and sciences was associated with low
achievement in analytical reasoning.

Such studies lend support to the notion that curricular programs can be
designed that enhance student learning. Consequently, as assessment information
from a test such as COMP is being used to assist institutions in the evaluation of
programs, faculty and administrators can be assured that improvements in programs
lead to improvements in student learning.
Assumption Two: Legitimate Use of Tests to Assess Programs

An additional assumption of this study is that tests can legitimately be used to assess the programs that have an impact upon student change. Traditionally, however, tests have been designed and used to measure traits possessed by individuals. In these cases, substantiation of test validity requires evidence that the content of the test adequately represents the trait being measured as well as evidence that the test is accurately measuring within the individual the trait that it claims to be measuring (Crocker & Algina, 1986). Tests, however, can be designed and used to assess programs, or, in other words, to measure and make inferences about traits possessed by groups of individuals (Petrie, 1987). In these cases, validity studies must provide different types of evidence than have been provided by traditional studies of tests designed for measuring traits within individuals.

The subject of test validity concerns the appropriateness of inferences or conclusions made on the basis of test scores (Messick, 1987). Current understandings of and methodologies for studying test validity are the result of an evolution of thought on the subject (Angoff, 1988). At various points throughout this evolution, different types of validity have been promoted. The major types have included: (a) predictive validity, in which a test is determined to be valid if it correctly predicts some future behavior; (b) concurrent validity, in which a test is determined to be valid if it correlates with another criterion of the same trait measured at the same time; (c) content validity, in which a test is determined to be valid if subject-matter experts verify that the items represent a satisfactory sampling of the subject domain; and (d) construct validity, in which a test is determined to be valid if theories about the nature of the construct measured by the test are confirmed (Cronbach & Meehl, 1955; Anastasi, 1968; Cronbach, 1971; Cronbach, 1984;
Messick, 1987). Today, less importance is given to the categories of validity, while it is largely accepted that various types of validity evidence best contribute to the meaningfulness of test scores. Most importantly, it is widely recognized that any conclusions about the validity of a test must be based on evidence that supports the use of the test results; validity belongs not to the test itself, but to the particular uses of the test (Cronbach, 1971; Messick, 1987).

The problems associated with using test results to make inferences about matters for which the test has not been proven to be valid, for example, using tests to evaluate programs that were in fact designed and validated to measure individuals, are evidenced in recent uses of SAT scores. The SAT was designed and validated for the purpose of assessing achievement levels of individual students preparing to pursue a postsecondary education (Donlon, 1984; Angoff, 1986). However, SAT scores have also been used to make judgments about the effectiveness or quality of education within a certain state or school district, a purpose for which the test was never intended nor validated.

Several researchers have pointed out the problems associated with using the SAT as a measure of educational effectiveness or quality (Powell & Steelman, 1987; Wainer, 1986; Linn, 1987; Gohmann, 1988; Fetler, 1991). These problems include: (a) lack of content validity as an assessment of secondary school instructional goals and content; (b) contamination of the research design associated with measuring covariates, such as average per pupil expenditures in a state or school district, on more students than actually took the SAT; (c) the confounding effects of program, school, or district with other variables that covary with SAT and that were omitted from the research model, such as student aptitude; (d) the selection bias that derives from the voluntary participation of SAT test-takers; (e) the differential selection bias that results from SAT participation varying from state-
to-state or school-to-school; and (f) construct validity problems associated with believing that SAT is a measure of educational quality. All of these problems suggest that the SAT has been used to make judgments about programs despite the fact that its purpose, administration practices, and validity evidence have had to do with the assessment of individuals rather than programs.

The COMP was one of the first tests designed specifically for the purpose of providing information to be used to assess programs. As such, questions have repeatedly been raised about the most appropriate ways to validate an instrument of this type. Researchers have noted that criteria for evaluating tests as measures of program effectiveness have not yet been clearly studied or established (Pike & Banta, 1989). At this point, three criteria seem reasonable to postulate; these criteria neatly address the problems cited above with respect to uses of the SAT:

1. An instrument used to assess programs must be content valid. For tests designed to assess individuals, content validity is based upon the extent to which the test items satisfactorily reflect the nature of the trait being measured. For tests designed to assess programs, the content of the test items must sufficiently reflect the goals and content of the curriculum or program being assessed.

2. An instrument used to assess programs must be validated upon its ability to be sensitive to program differences. For tests designed to assess individuals, validity evidence is based upon the extent to which the test results accurately reflect differences among individuals in the amount of the specific trait being measured; such evidence may in part be gathered by correlating the test scores with those of another established test of the same trait. For tests designed to assess programs, validity evidence must be based on the extent to which the test can differentiate among programs that are likely to produce different test results; such evidence can in part be gathered by correlating the test scores with variables that represent
differences in program structure. Appropriately measured covariates should be considered for use in the validity research design.

3. Validity evidence depends upon the appropriateness of the test administration practices used to gather validity data. Tests designed to assess programs should be administered either to random samples of program participants or to the entire population of program participants. This enables researchers to draw reliable and valid inferences about the effects of the program.

Given that these criteria are met, it is reasonable to assume that tests such as COMP can be used to make legitimate judgments about the effectiveness of programs.

Validity of COMP as a Measure of Program Effectiveness

Numerous validity studies of COMP have been conducted by the ACT company as well as by independent researchers and users of the instruments. Nearly all of the studies of COMP can be grouped within one of two general types: (1) content validity studies in which faculty members or panels of experts are asked to examine the match between COMP objectives or individual items and a particular college's general education goals and course content; and (2) correlational studies that look at associations between COMP scores and other individual student characteristics such as gender, age, type of coursework completed, and GPAs. Results of notable studies from each of these two areas will be presented in this section. In addition, there are four studies in the literature that in some way attempt to examine the effects of variation in general education programming upon COMP scores; these too will be discussed in some detail. By way of introduction, however, an overview of the process used to develop COMP and a description of the structure of the instrument are first presented.
Development and Description of the COMP Instrument

Persons who developed COMP (Forrest & Steele, 1982) have reported that the test development process first included: (a) reviews of the existing theoretical and empirical literature on general education; (b) solicited input from a large number of distinguished educators including state department executives, staff members of accrediting agencies, and college and university faculty and administrators; and (c) group decisions about how to define the cognitive components of the construct general education and the processes and content knowledge these components require. As a result of this preliminary work, COMP was specifically designed to assess three content and three process outcome areas. The content outcome areas are Functioning Within Social Institutions, Using Science and Technology, and Using the Arts. The process outcome areas are Communicating, Solving Problems, and Clarifying Values. Figure 1 describes these six outcome areas in detail.

Viewed as being independent of specific college courses yet common to the overall goals of most college and university general education programs, these six content and process outcome areas provide the theoretical construct definition of general education for the COMP instruments. While the specifics of general education program content and delivery may differ from institution to institution, it is this commonality of goals that enables COMP to be potentially useful for general education assessment at virtually any college or university and allows users of the test to compare their own scores with those of other institutions.

Over the years, COMP has included several different instruments. The initial assessment materials developed in the 1970s as part of COMP were assembled into an instrument referred to as the Measurement Battery. This Battery
Content Areas

**Functioning Within Social Institutions:** Can identify those activities and institutions which constitute the social aspects of a culture (for example, governmental and economic systems, religion, marital and familial institutions, employment and civic volunteer and recreational organizations); understand the impact that social institutions have on individuals in a culture; analyze one's own and others' personal functioning within social institutions.

**Using Science and Technology:** Can identify those activities and products which constitute the scientific/technological aspects of a culture (for example, transportation, housing, energy, processed food, clothing, health maintenance, entertainment and recreation, mood-altering drugs, national defense, communication, and data processing); understand the impact of such activities and products on the individuals and the physical environment in a culture; analyze the uses of technological products in a culture and one's personal use of such products.

**Using the Arts:** Can identify those activities and products which constitute the artistic aspects of a culture (for example, graphic arts, music, drama, literature, dance, sculpture, film, architecture); understand the impact that art, in its various forms, has on individuals in a culture; analyze uses of works of art within a culture and one's personal use of art.

Process Areas

**Communicating:** Can send and receive information in a variety of modes (written, graphic, oral, numeric, and symbolic), within a variety of settings (one-to-one, in small and large groups), and for a variety of purposes (for example, to inform, to understand, to persuade, and to analyze).

**Solving Problems:** Can analyze a variety of problems (for example, scientific, social, artistic, personal), select or create solutions to problems, and implement solutions.

**Clarifying Values:** Can identify one's personal values and the personal values of other individuals, understand how personal values develop, and analyze the implications of decisions made on the basis of personally held values.

---

Figure 1. The Six Outcome Areas Measured by COMP (American College Testing Program, 1992).


Used with permission of Joe M. Steele, Ph.D., Director, College Outcome Measures Program, 12-17-93.
included fifteen assessment activities that, taken together, sampled abilities in all six of the areas indicated in Figure 1. Administration time for the entire Battery was six-and-a-half hours. Stimulus materials for the assessment activities included videotapes, audiotapes, as well as written narratives; all responses were collected in an open-response format that included short answers, essays, and taped oral responses.

In 1978, the Composite Examination replaced the Measurement Battery as the operational mode for COMP and is still available for use today. Like the Battery, the Examination continues to use the series of fifteen activities that sample abilities in all six areas. Also like the Battery, some items on the Examination call for open written or spoken responses, including some lengthy expository passages. Unlike the Battery, some items on the Examination require that examinees select from among several possible answers in a multiple-choice format. The Examination requires four-and-a-half hours of testing time.

Development of the Objective Test was begun in 1978 as a shorter "proxy" measure for the Examination. The Test requires two-and-a-half hours of testing time and the item format is entirely multiple-choice. Like the Examination and the Battery before it, stimulus materials for the individual items are not entirely in written format; materials continue to include videotapes and audiotapes, as well as written narratives. The Test offers colleges and universities a less expensive way of testing large groups of students in order to conduct evaluations of general education programs.

One final point needs to be made regarding the structure of COMP; the scoring protocol is such that each test item is utilized to measure a process outcome in the context of a particular content outcome. As a result, each test item contributes information to two subscale scores, one being a content subscale and the
other a process subscale. More will be said about this in the discussions about methodology; COMP's scoring protocol has important implications for any analysis of its scores.

Studies of the Validity of COMP

As noted above, nearly all of the studies of COMP can be grouped within one of two general types: (1) content validity studies in which faculty members or panels of experts are asked to examine the match between COMP objectives or individual test items and a particular college's general education goals and course content; and (2) correlational studies that look at associations between COMP scores and other individual student characteristics such as gender, age, type of coursework completed, and GPAs. The content validity studies are extremely important in that one of the criteria postulated as being important for tests of program effectiveness is that the test be content valid. The correlational studies, however, provide little substantial contribution to the remaining COMP validity issue: the extent to which COMP scores are explained by differences among programs. In fact, these methodologies have been adopted from those used to study tests designed for assessment of individuals and contribute little information to the validity of COMP as a measure of programs.

The following sections explore the content validity studies and highlight the findings of the correlational studies. In addition, four studies will be examined that specifically address COMP's ability to be sensitive to differences among general education programs.

Studies of Content Validity

McGuire (1987) asked faculty from disciplines that provided the general
education component at a moderately large midwestern community college to form a content review panel for purposes of rating each item on the COMP instrument in terms of: (a) students' exposure to the content of the item in general education courses, and (b) importance of the outcome in terms of students' general education. Faculty rated COMP highly in terms of exposure; 92% of the test items fell below the midpoint of the exposure scale, where lower scores indicated greater likelihood of exposure to the test material. The skills and knowledge tested by the items were also judged to be important, with 100% of the items falling below the midpoint of the importance scale, where lower scores indicated greater importance.

An evaluation committee at the University of Montevallo (McLean et al., 1989) conducted a similar study, recognizing that in order to demonstrate that COMP is content valid, three components must converge: (1) the knowledge and skills being measured by COMP, (2) the goals of Montevallo's general education program, and (3) the content of Montevallo's general education courses. To carry out their study, sixteen faculty members at Montevallo, all involved in teaching general education courses, were selected to form a review panel that examined COMP items. Panel members were specifically asked to respond to three questions: (1) were there any general education goals at Montevallo for which there were no corresponding COMP goals; (2) were there any general education goals that appeared subtly different from the apparently equivalent COMP goals; and (3) were there any test questions which they felt to be misleading, incorrectly keyed, or not really testing the general education and/or COMP goal which they were supposed to address. In addition to the use of this panel, the content validity of the test was further addressed by surveying all faculty who taught the general education curriculum courses (Rogers & McLean, 1986). Results from the panel and the survey led the authors to conclude that COMP adequately assessed most of
Montevallo’s general education goals.

In a similar study, Pike and Banta (1989) evaluated the content coverage of COMP using a panel of faculty to compare the content of COMP with the goals for general education at the University of Tennessee at Knoxville (UTK). Panel members there felt that the correspondence between COMP and UTK’s general education goals was highest for UTK’s goals in the areas of solving problems and reading skills (extensive coverage of these goals by COMP items), followed by aesthetics (strong coverage), followed by technology, values, and political dynamics (moderate coverage), followed by spoken English, computational skills, science for life, economics, and social sciences (fair coverage). Members of the faculty panel concluded that content coverage was less than hoped for, but that COMP could be used to assess general education outcomes as long as its limitations were recognized.

Colleges and universities generally conduct their own content validity studies prior to adopting the COMP for use in assessment programs; results of such studies are generally not placed in the public domain. The fact that several hundred colleges and universities have used COMP attests to the fact that many institutions find the goals and content of COMP to be compatible with their own general education goals and content. This fact validates the theoretical assumption that, in the judgment of faculty and administrator panels, COMP outcomes are those held in common by nearly all college and university general education programs.

Validity Studies Involving Correlational Designs

Much of the validity work that has been done on COMP has involved various types of correlational methodologies. Researchers have correlated individual COMP scores with: (a) student demographic characteristics, (b) pre-
college student achievement as measured by the ACT Composite, (c) length of
exposure to higher education as measured by number of semesters in college, (d)
amount and type of coursework completed, and (e) student GPAs. These studies
have attempted to determine the nature of the construct measured by COMP,
particularly as it might be associated with relevant individual student characteristics.
Following are several of the more important conclusions that have come from this
research:

1. No strong associations have been found between student demographic
characteristics (gender, ethnic origin, and age) and COMP scores \( r = .01 \) to \( r = .11 \),
thus lending support to the premise that COMP is a measure, not of student
demographic differences, but, perhaps, of what it is intended to measure -- general
education knowledge and skills learned during college (Forrest & Steele, 1982;
Pike, 1989a).

2. Moderate to strong associations \( r = .60 \) to \( r = .86 \) have been found
between COMP scores and measures of pre-college student achievement such as the
ACT Composite or the SAT (Forrest & Steele, 1982; Pike, 1989a). Pike (1989b)
found pre-college student ability to be the primary determinant of performance on
COMP; ACT Composite scores accounted for 37% of the variance in COMP Total
scores. It is important to note that the percent of variance explained by a particular
variable in question is clearly related to the nature of the full variable set used. Still,
these findings have led some researchers to conclude that COMP may be more a
measure of student ability than of what has been learned in college, thereby
questioning the validity of COMP as a measure of program effectiveness.

3. Positive correlations have been found between COMP scores and other
measures of general education, such as the Educational Testing Services' Undergraduate Assessment Program Area Tests \( r = .54 \) to \( r = .69 \) and the
Academic Profile ($r = .15$ to $r = .57$) (Forrest & Steele, 1982; Pike & Banta, 1989). In general, the evidence suggests that COMP scores appear to correlate with other measures of similar constructs.

4. Several studies have shown low to moderate but significant ($p < .05$) correlations between COMP subscale scores and the academic majors or coursework patterns completed by students ($r = .01$ to $r = .23$) (Forrest & Steele, 1982; Fike, 1989a; Pike & Banta, 1989; Pike & Phillippi, 1989). In these studies, however, sample sizes tended to be quite large (roughly 6,000 students), a factor that certainly contributed to the statistical significance of the rather meager correlations. In general, however, larger correlations were found between majors or coursework patterns and subscale scores than between majors or coursework patterns and COMP Total score. This lends support to the validity of the COMP battery as a measure of general learning and the subscales as measures of more specific content learning.

6. Moderate correlations ($r = .21$ to $r = .46$) have been found between COMP scores and the more traditional measure of college learning—the GPA (Forrest & Steele, 1982; Schomberg et al., 1982; Sibert & Ayers, 1989; Olsen, 1990). The expectation is that these two measures would be associated, since GPA is an indication, although not standardized across institutions, of student performance and learning.

Validity Studies of COMP That Address Differences in General Education Programs

None of the studies discussed thus far speak to the second of the two primary issues involved in examining the validity of COMP as an instrument used in general education program assessment: the effect of different types of general
education programs upon students' learning of general education knowledge and skills. There are, however, four studies that stand out from those already discussed; these four studies, discussed below, come closest of any currently in the literature to examining the effects of variation in general education programming upon COMP scores. Each, however, has some weakness that suggests that further research in this area is needed.

The first study, by Forrest and Steele (1982), attempted to isolate the effects of a general education program by comparing students who had been exposed to little or no general education component (vocational-technical students) with those who had been exposed to a full general education component as part of their four-year degree program. Fifty (50) randomly selected college seniors were matched with 50 vocational-technical students on entering ACT Composite scores, age, and gender. The mean COMP Total score was substantially higher (mean = 44.52) for the baccalaureate graduates than for the graduates of the vocational, two-year program (mean = 27.22). (This study used a modified version of COMP, and the scores obtained are not comparable to those obtained in other COMP studies. The national mean on COMP is actually around 180.) All six subscale scores were also substantially higher for the graduates of the four-year program. Since the two groups were judged to have the same level of entering college achievement, a tentative conclusion was that the general education experience of the college seniors had a greater impact upon their knowledge and skills than did maturation or life experience on the part of the vocational-technical students. This is a particularly important finding, since it lends significant support to the validity of COMP as a measure of general education knowledge and skills gained in college programs. What is needed to complement this study is an examination of the relationship of COMP scores to different types of general education programs.
The second study was the only one to utilize data from several different colleges and universities. Forrest (1982) combined data from 44 institutions that had administered the COMP Examination to representative groups of sophomores or seniors for whom pre-college ACT Composite or SAT scores were also available. Included were five liberal arts colleges that are units of large public research universities, five public community colleges, eight small private colleges that are above average in wealth and admission selectivity, six small public colleges, eight small private colleges below average in wealth and admission selectivity, five small private universities, and seven large public comprehensive universities.

Fifty (50) different institutional characteristics were looked at in terms of their relationship to institutional outcomes, of which COMP scores were one measure. The institutional characteristics fell into five broad categories: (1) orientation and academic advising of new students, (2) general education curricular objectives and requirements, (3) instructional dimensions, (4) extracurricular activities, and (5) general institutional characteristics. Three of Forrest's findings are particularly important in terms of the evidence they provide about the validity of COMP:

1. The amount and breadth of general education coursework required of students appears to be related to COMP gain scores. Forrest looked at the general education components of two groups of the 44 institutions differentiated in part by the size and style of course distributions within the general education program. One group of eight institutions allotted an average of 46% of the baccalaureate degree to general education and distributed the general education requirements fairly equally among four broad areas: (1) communications, (2) social science and history; (3) natural science and mathematics; and (4) fine arts and humanities. The other group of nine institutions allotted an average of 31% of the baccalaureate program to
general education and did not attempt to provide equal distribution of requirements into these four categories. Controlling for pre-college ACT Composite scores, the mean COMP score gain of the first group was 8.9 raw gain-score points, while for the second group the gain was only 3.8 points.

2. The existence of general education goal statements and the use of proficiency examinations seems to be related to COMP gain scores. Of the 44 institutions included in the study, five were identified as having both student-oriented general education goal statements and proficiency examinations while 18 institutions were identified as having neither. For the first group, the average COMP Total score gain was 11.6 raw gain-score points, and, for the second group, the average score gain was 5.9 points.

3. An institution's ability to individualize instruction and focus it on relevant, practical skill building within the structure of the formal general education program appears to be related to COMP gain scores. Looking at instructional dimensions, 19 institutions were identified as having both formal remedial programs and formal, credit-bearing, off-campus general educational opportunities such as field observation, internships, and on-the-job experiences; 13 institutions were identified that had neither. Average COMP score gain was 8.8 raw gain-score points for the first group and 4.6 points for the second.

The strength of this study was its attempt to examine the performance of COMP within the context of multiple institutions having several different types of general education programs. In addition, the study went beyond the structural features of general education curriculum and looked at COMP's sensitivity to various other academic support programs such as advising, individualized and remedial instruction, and field experience or internship programs. This study found some interesting and important associations between COMP gain scores and several
program elements.

This study also had several weaknesses. First, the focus of the different analyses was usually between clusters of institutions that were on the extremes of the variable being examined. For example, gain scores were compared between two clusters of institutions, one having the most comprehensive advising and orientation programs and the other having the least comprehensive advising and orientation programs. This approach, which amounted to eliminating the full range of program variation, artificially maximized between-group differences. Second, although numerous programs and institutional variables were examined, the analyses were generally bivariate; COMP gain scores were examined in relation to one type of program variable at a time. Therefore, other institutional concomitant variables might have been unknowingly confounding the reported relationship between program features and COMP gain scores. Third, while raw gain-score point differences were reported, neither statistics resulting from tests of significant differences nor computation of correlations were reported. The size and significance of the associations that were found are difficult to interpret.

The third study of this type was conducted at the University of Minnesota, one of 50 colleges and universities selected in 1979 to participate in a field study to test the administration, validity, and reliability of COMP (Schomberg et al., 1982). The important question addressed by the study was: can COMP differentiate between the different types of baccalaureate programs offered by the University of Minnesota. The study involved testing seniors who were enrolled in the College of Liberal Arts (CLA), the General College, the Inter-College Program (ICP), or the University Without Walls (UWW); seniors were defined as students having completed 160 or more credits or having entered the graduation phase of the University Without Walls program. The College of Liberal Arts, the most
traditional of the four programs, was characterized by the successful completion of requirements in the areas of the discipline-based major, general education distribution requirements, and freely-chosen electives; nearly all learning experiences were classroom based. The General College baccalaureate was individually designed and allowed for substantial amounts of vocationally-oriented activities to be included as learning experiences. The Inter-College Program was also individually designed, combined courses from different colleges, and had a strong career orientation. The University Without Walls baccalaureate was not credit based, but there were liberal arts education criteria and students completed a major concentration; students were encouraged to integrate experiential learning along with the completion of actual courses, often in the form of independent study. The researchers hypothesized that COMP scores would differentiate among seniors of these different baccalaureate programs.

The results of the study indicated that COMP did detect some gross differences among subgroups of students who both had different curricular emphases and length of exposure to higher education, but did not differentiate among groups of students who had somewhat different curricular experiences but had the same length of exposure to higher education.

Although the general design for this study was quite appropriate, particularly for university faculty and administrators wishing to evaluate the comparative quality of somewhat nontraditional baccalaureate programs with traditional degree programs, there were some unfortunate methodological problems that may have caused spurious results. One problem was the small sample size; as few as five students represented some of the baccalaureate programs. In addition, students who sat for the test were all volunteers who were, perhaps, most interested in the incentive of receiving a one-dollar pizza coupon.
The fourth study of this type, by Steele (1989), has shown that colleges that have made changes in their general education programs based upon COMP results have found that such changes do affect future COMP results. Drawing from baseline cross-sectional and longitudinal data, Bethany Lutheran College in Mankato, Minnesota made several changes in their general education curriculum, including incorporating more practical applications of skills, experiential and off-campus learning experiences, and student involvement in doing investigative research. These changes resulted in substantial changes in future COMP scores, particularly on the subscales Using Science and Technology, Communicating, and Clarifying Values. Our Lady of the Lake University in San Antonio, Texas, began to use COMP scores to screen entering students; those students not achieving the minimum levels of proficiency determined by the university received further instruction in relevant areas and were retested. Steele reported that significant gains were made by these students and that faculty attributed the gains to the instruction given. Lastly, the College of Nursing at Austin Peay State University in Clarksville, Tennessee used information from COMP scores to develop instruction in the areas of Clarifying Values and Communicating. Substantial increases were found in these subscale scores in subsequent testings. Although the results of these case studies are very interesting, the methodology was certainly not rigorous enough to allow one to infer that changes in general education clearly led to higher COMP scores.

Conceptual Framework for Further Studies of the Validity of COMP

The COMP represents a new development in measurement science: the design and use of tests specifically for purposes of analyzing group data and making decisions about program effectiveness. Given the current interest in assessment, institutional effectiveness, and total quality management, COMP surely will not be
the last test developed for this purpose. In fact, the Educational Testing Service began development of the Academic Profile in 1986, a test that has also been designed to give institutions some means of assessing the outcomes of their general education programs. After three years of field testing, the instrument is now fully operational for use by colleges and universities (Educational Testing Service, 1990).

An important point made earlier in this chapter is that methods that have been used in the past to validate instruments designed to assess individuals are not sufficient for determining the validity of instruments designed to evaluate programs. A discussion was included above of the problems associated with using the SAT, a test designed and validated for the assessment of individuals, for drawing inferences about program effectiveness. Researchers need to reexamine their methodologies and explore new ways to validate instruments such as COMP and the Academic Profile, instruments intended for program assessment.

The last four studies described in the above section represent the most promising conceived approaches thus far to doing validity research of instruments designed to assess the effectiveness of programs. Each attempted to examine the effects of different programs or changes in programs upon COMP scores. What is still needed, however, are inter-institutional examinations of the effects of differences among general education programs upon COMP scores while considering the possible covariation of multiple institutional variables. In addition, consideration must be given to the full range of program variation. Only by knowing that COMP scores are sensitive to program differences can it be certain that COMP is valid for use as a measure of general education program effectiveness.

The purpose of this study is specifically to determine whether COMP scores are meaningfully explained by general education program variation. The next sections will describe a proposed conceptual framework for conducting this type of
study.

**Defining the General Education Program**

General education programs are based on the notion that a college education should in part provide students with a common core of knowledge, skills, and values (Boyer, 1981; Gaff, 1983; Miller, 1988; Purves, 1988). The fact that the specific characteristics of general education programs in our colleges and universities are vast and varied points out that educators have not been able to agree on what the common core should contain. Some of the ways in which college and university general education programs differ include: (a) the number of courses required or the percentage of total undergraduate coursework that is given to general education; (b) the disciplines represented in the course requirements; (c) the breadth versus the depth of general knowledge that students are expected to achieve; (d) the amount of control that institutions assume over the courses that students may choose; (e) whether or not the approach to general education is interdisciplinary or subject specific; (f) the way in which course requirements are distributed throughout the four years of an undergraduate program; (g) the modes of instructional delivery; and (h) whether course goals include explicit attempts to build thought processes, such as critical thinking, in addition to building content knowledge (Blackman, 1964; Rudolph, 1977; Gaff, 1983).

The various elements used to differentiate general education programs can be organized into two broad categories: (1) elements that are institutional-structural, and (2) elements that are student-development oriented (Amey, 1992). Another way of referring to these divisions would be to speak of the formal curriculum, that is, the one described in the college catalog, and the informal curriculum, that is, the one that is actually experienced by students going through the curriculum. Structural
elements are those things that provide the overarching framework such as written goals and objectives, degree requirements, course requirements, course content, size of classes, and regulations regarding the curriculum. Elements that are student-development oriented are instructional methodologies, nature of interaction between faculty and students, explicit or implicit incentives for learning, and values, attitudes, and beliefs that are transmitted through the instructional process. A major difference between these two categorizations of the curriculum is that the formal elements typically receive far more attention in terms of planning, analysis, documentation, and administrative control. Both categories, however, constitute valid approaches to studying the general education curriculum.

A thorough examination of the ability of COMP to be sensitive to program differences would necessitate studying the effects of both categories of curriculum elements. However, for two reasons, this particular study will focus on the institutional-structural elements of general education programs. First, the areas of curricular organization that have received the most attention from institutions that have worked to reshape their general education curriculum have involved the number and type of courses required and the freedom of choice given to students over their courses (Gaff, 1983). Second, the focus of both state and federal government general education recommendations or mandates has been on structural components. For example, in 1989, the National Endowment for the Humanities released its report, *50 Hours: A Core Curriculum for College Students*, recommending the specific number and content of courses for a general education curriculum (Cheney, 1989). That same year, the Massachusetts Board of Regents of Higher Education called on all state colleges and universities to establish a core curriculum that would include certain thinking skills and prescribed areas of study (Ingalls, 1989). So, an important question that must be answered is whether or not
COMP is able to provide the type of information that higher education and government leaders are seeking as they hope to improve general education through such restructuring.

Two institutional-structural curriculum elements will be the particular focus of this study: (1) the percentage of total undergraduate coursework that is devoted to general education; and (2) the amount of control that an institution assumes over students' selection of general education courses. Together these two elements provide a comparative framework for general education that colleges and universities across the country can understand and appreciate. These elements cut to the quick of what is meant by curricular structure and coherence, breadth and depth. Together these elements speak to the principle recommendation of Integrity in the College Curriculum, that "the curriculum requires structure, a framework sturdier than simply a major and general distribution requirements and more reliable than student interest" (Association of American Colleges, 1985, p. 15).

Additional Variables That Affect Learning

A study of the relationship between particular features of the general education curriculum and COMP scores would be invalid if it did not consider additional variables that have been found to correlate with post-college student achievement. Such variables, for example, measures of pre-college achievement, faculty-student interaction, and quality of student life, must be included in conceptual and analytical models in order that the effects of these variables upon student learning can be separated from the effects of the general education program. To do otherwise would be to invite the high probability of confounding the effect of general education with the effects of other variables impacting student learning.

Although not entirely conclusive, the higher education literature points to
several concomitant variables that should be considered if one is trying to determine the effects of a particular program upon learning outcomes. For purposes of this study, the following variables will be considered:

1. Pre-College Student Achievement. There is strong, conclusive evidence that pre-college achievement, measured by the ACT Composite, SAT, or National Merit Scholarship Qualifying Test (NMSQT), is a strong predictor of post-college achievement, measured by the GRE or COMP tests (Nichols, 1964; Astin, 1968; Rock et al., 1970; Forrest & Steele, 1982; Pike, 1989a; Astin, 1993).

2. Proportion of Students Who are Transfers From Another Institution. There are slight indications that transfer behavior, particularly from one four-year institution to another, tends to have a negative influence on educational attainment (Pascarella, 1985).

3. Proportion of Students Living On Campus or in Social Houses. Evidence suggests that students living on campus or in social houses (fraternities and sororities) may be at a disadvantage in terms of their cognitive development during college (Winter, McClelland, & Stewart, 1981).

4. Interaction of Faculty and Students. The amount of out-of-class and informal interaction between students and faculty appears to be a strong, positive factor in the academic achievement and aspirations of students (Astin, 1968; Centra & Rock, 1971; Wilson, Gaff, Dienst, Wood & Bavry, 1975; Pascarella & Terenzini, 1978; Terenzini & Pascarella, 1980). In addition, it seems that possibilities for greater interaction exist where the ratio of students to faculty is low (Pascarella & Terenzini, 1991).

5. Proportion of an Institution's Faculty with Doctorates. A few studies have provided evidence suggesting that post-college student achievement, measured by GRE while controlling for pre-college SAT scores, is significantly associated
with the proportion of an institution's faculty holding a doctorate (Rock et al., 1970; Ayres & Bennett, 1983; Astin, 1993).

These five variables, evidenced to be associated in varying degrees with students' post-college achievement levels, should arguably be considered in a study looking at the effects of general education program structure upon COMP scores.

Studying the Validity of COMP

The conceptual model being suggested here as one way to study the validity of COMP includes the following methodological components: (1) uses comparative data from multiple institutions which thus represent diverse general education programs; (2) focuses on the institutional-structural elements of general education that are common to all programs and that have typically been the subject for debate and restructuring during times of reform; (3) considers student inputs, that is, the relationship between students' pre-college achievement level and their post-college achievement level; and (4) considers environmental effects, that is, the effects of certain institutional characteristics other than those specific to the general education program upon students' post-college achievement. This model very simply allows the association between COMP scores and general education programs to be examined while controlling for possible effects of other college experiences that might also result in students' increased knowledge and skill in the areas of general education. Although fairly simple in design, this model has not yet been adequately applied to a study of COMP.

This model, along with its analytical analog, will be discussed in greater detail in the following chapter.
CHAPTER III

STUDY DESIGN AND METHODOLOGY

The specific problem addressed by this study concerns the sensitivity of COMP to differences among college and university general education programs. Despite the importance of this issue, there is general agreement that such an effect has not been adequately researched (Yarbrough, 1992). Therefore, the purpose of this study is to contribute evidence that either supports or refutes the validity of COMP by determining whether or not the test is able to distinguish among particular structural differences of general education programs. The purpose of this third chapter is to present the conceptual model driving the research design as well as to describe the data collection and analytical procedures used in the study.

The Conceptual Model

As pointed out in earlier chapters, COMP was specifically designed to be used to evaluate a college or university general education curriculum by assessing the knowledge and skills attained by students as a result of completing such programs. The specific concern of this study is with whether or not COMP scores are meaningfully explained by differences in general education programs while considering the effects of other institutional variables that might also be affecting student achievement.

The conceptual model for this study is presented in Figure 2. The outcome in this model is post-college general education achievement of an institution's students as measured by COMP. An important point must be made about this
outcome or endogenous variable. In most studies, the central issue would be the construct, namely achievement, and the instrument used to measure achievement would serve to operationalize the construct but, in a sense, would be of secondary importance. Since the problem specific to this study is the validity of COMP and not achievement, the complete construct or outcome under examination is post-college general education achievement as measured by COMP.

Two constructs are expected to impact upon post-college achievement as measured by COMP, thus serving as the explanatory variables in the model: (1) pre-college achievement of the institution's students, and (2) the students' experiences of college. For purposes of this study, the latter construct is divided
into two areas: (1) characteristics of the general education program, and (2) characteristics of the institution that are known in some way to affect learning.

The association that is central to this study is that between the general education program characteristics as independent variables and institutional COMP scores as dependent variables. In the diagram of the conceptual model, an arrowed straight line connects these two variables indicating that this is the key relationship under study. As discussed in earlier chapters, this relationship must be looked at in the presence of other variables that have been found to affect college learning. Therefore, pre-college achievement and characteristics of the institution also serve as explanatory variables, more specifically, concomitant variables: variables that are in attendance but are not of prime importance. The task of the analysis is to statistically control for their effects upon COMP so that the effects of the general education program might be isolated and examined. In the diagram of the model, dotted lines connect these concomitant variables with COMP to distinguish them from the importance of the general education program in this study.

Design of the Study

From the conceptual model presented above, it is clear that the general hypothesis of this study is that differences among institutions in terms of the structure of their general education programs are related to institutions' COMP scores and that this relationship provides evidence that supports the validity of COMP. Stated more specifically, the two types of hypotheses tested in this research were:

1. There is a relationship between an institution's COMP Total or subscale scores and the amount of control that an institution assumes over students' selection of their general education courses, measured as the ratio of number of general
education courses required to the number of courses from which students may choose when completing their general education requirements.

2. There is a relationship between an institution's COMP Total or subscale scores and the number of general education courses that students take, measured as the percent of total undergraduate coursework that is devoted to general education.

In the detailed discussion of the study design that follows, five specific issues will be addressed: (1) the dependent variables, (2) the independent variables, (3) the concomitant variables, (4) the definition of the study population and sample, and (5) the level of the data to be used in the analysis. Of great importance is the fact that any research study involves certain trade-offs in design and methodology. The goal is to conduct a study that enables the researcher to draw valid inferences and make fair generalizations to the study population while being cognizant of the flaws that necessarily remain. One principal aim of this chapter is to discuss these trade-offs and to provide the rationale for the choices that were made in designing this study.

Variables

A complete discussion of the study variables includes reference to the dependent, independent, and concomitant variables. Throughout this section, the reader may find it helpful to refer to Table 1 for a listing the study variables.

Dependent Variables

The dependent variables for this study are COMP results (institutional means), reported by the ACT company in a Total score as well as subscale scores for three content outcome areas and three process outcome areas. The three
<table>
<thead>
<tr>
<th>Type of Variable</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variables:</strong></td>
<td></td>
</tr>
<tr>
<td>Functioning Within Social Institutions Subscale Means (FSI)</td>
<td></td>
</tr>
<tr>
<td>Using Science and Technology Subscale Means (US)</td>
<td></td>
</tr>
<tr>
<td>Using the Arts Subscale Means (UA)</td>
<td></td>
</tr>
<tr>
<td>Communicating Subscale Means (COM)</td>
<td></td>
</tr>
<tr>
<td>Solving Problems Subscale Means (SP)</td>
<td></td>
</tr>
<tr>
<td>Clarifying Values Subscale Means (CV)</td>
<td></td>
</tr>
<tr>
<td><strong>Independent Variables:</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Concomitant Variables:</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

content area subscales are Functioning Within Social Institutions, Using Science and Technology, and Using the Arts; the three process area subscales are...
Communicating, Solving Problems, and Clarifying Values. Detailed descriptions of these outcome areas were provided in Figure 1 in Chapter II.

As mentioned in an earlier chapter, the design of the scoring procedures for COMP are such that each test item is measuring a process outcome in the context of a particular content outcome. For example, the subscale Functioning Within Social Institutions is measured by items that require the test-taker to communicate about social institutions (also measuring Communicating), solve social problems (also measuring Solving Problems), and clarify social values (also measuring Clarifying Values). As a result, each test item contributes information to two subscale scores, one being a content subscale and the other a process subscale.

The subscales within each outcome area, although moderately correlated, are independent measures; for example, the items contributing to the Functioning Within Social Institutions subscale score are distinct from the items used to compute the other two content subscale scores—Using Science and Technology and Using the Arts. The subscales crossing the outcome areas, however, are not independent measures; for example, some of the same items are used to compute the scores for both the Using the Arts subscale (a content area) and the Clarifying Values subscale (a process area). This dependence among scales has long been recognized as problematic in the interpretation of test battery results for individuals; in fact, it is considered irresponsible to make inferences about supposedly distinct traits from test scales that are highly dependent (Cooley, 1971). In like manner, to use COMP scores to make inferences about a program's ability to affect students' use of the arts and problem-solving skills without regard for the structural dependence of the two scales is highly problematic.

One practical complication resulting from this scale dependence is that institutions working to raise scores on content subscales will also implicitly raise
process subscale scores without specifically attending to that process outcome; the traits that cross outcome areas cannot be separated. An analytical complication is that subscale scores that cross outcome areas cannot be incorporated into the same statistical models; for example, to include scores from both Using Science and Technology and Solving Problems as explanatory variables in a single regression model would invite problems associated with multicollinearity and lack of independence among measures. Extreme caution must be taken in the analysis and interpretation of the COMP subscale scores.

That being said, it remains a fact that the ACT company provides test results in the form of a Total score and six subscale scores. Certainly, despite the problems associated with the dependence of the subscales, institutions in some way use all six of the subscale scores presented in the institutional test results profile. Therefore, a study focusing just on the Total score or just on some of the six subscale scores would provide incomplete information about the test to users of the test.

For this study, then, the COMP Total score and each of the six subscale scores will be used in turn as the dependent variable in separate analyses in order to determine whether or not these scores are sensitive to structural differences among general education programs. In no instance will subscales from different outcome areas be included in the same analytical model.

**Independent Variables**

The independent variables in this study are two particular structural components of general education programs: (1) the amount of control that an institution assumes over students' selection of general education courses, and (2) the percentage of total undergraduate coursework that is devoted to general education.

The first component of general education is operationalized by dividing, for
each institution included in the study sample, the number of courses required in
general education by the number of courses that students may choose from when
deciding the courses they will take. For ease of understanding, the resulting number
is multiplied by 100 and rounded to single decimal precision. For example, for a
college that requires six general education courses to be completed while allowing
students to choose from a selection of 18 courses, the variable value will be 33.3;
for a college that requires 10 general education courses to be completed but that
allows students to choose from a selection of 250 courses, the variable value will be
4.0. Higher values indicate greater control on the part of the institution over the
courses that students may take to complete their general education requirements.
This variable is treated as continuous rather than categorical.

The second component of general education is operationalized by dividing,
for each institution included in the study sample, the number of courses required in
general education by the total number of courses required to graduate with a
bachelor's degree. Again, the resulting number is multiplied by 100 and rounded to
single decimal precision. This variable is also treated as continuous rather than
categorical.

Concomitant Variables

The most valid approach for inferring clear, perhaps even causal, links
between dependent and independent variables is through the use of random
assignment of randomly-selected subjects to treatment groups. This type of
experimental design enables the researcher to put aside most concerns about
systematic variations that may confound the attribution of cause to the independent
variable (Cook & Campbell, 1979).

Unfortunately, in much educational and social research, randomization is not
possible. In this particular study, for example, it is not possible to randomly assign students to institutions having different general education programs to determine the effect of these differences on students' learning or on differences in COMP scores. Therefore, this research is necessarily quasi-experimental (Cook & Campbell, 1979), and there is a need to be concerned about the confounding effects of variables not incorporated into the conceptual model or design that can threaten the internal validity of the results, or, in this case, threaten the inferences made about the relationship between the general education program and COMP scores. Specifically, in this study one must be concerned about the many variables besides the structure of the general education program that possibly influence students' learning and, in turn, students' scores on COMP.

In quasi-experimental designs, variables that are known to covary with the dependent variable are wisely included in the design to control for differential input among subjects. In this way, the use of such concomitant variables introduces an adjustment of the dependent variable such that the effects of the covariation are removed. These adjusted response values (most likely, adjusted means) are then used to determine if the independent variable is significantly related to the dependent variable. This type of analysis of covariance essentially attempts to equalize the subjects on the basis of one or more concomitant variables before determining the effect of the independent variable. In regression analysis, this type of control is achieved by entering the concomitant variables into the equation prior to entering the independent variable(s) that is of primary interest.

As shown in the previous chapter, the educational literature suggests several variables that are likely to covary with measures of post-college achievement such as COMP scores. The effects of these variables can be statistically controlled so that inferences about the association between general education programs and
COMP scores may be more valid. First presented in Chapter II the following five variables were chosen to be included in this study as concomitant variables: (1) pre-college student achievement as measured by the ACT Composite test, (2) percent of students who are transfers from another institution, (3) percent of students who live on campus, (4) number of students per faculty member, and (5) percent of an institution's faculty with doctorates. For each variable, values are rounded to single decimal precision; percentages are the results of ratios first multiplied by 100.

Level of the Data

The level of analysis for this study was the institution. The decision to use institution-level data, however, presents certain problems. It is well known that correlations based on averages or rates, called ecological correlations, can be inflated and misleading (Robinson, 1950; Borgatta & Jackson, 1980). Correlations of means essentially ignore the variation of values around the means and give a misleading impression of tight clustering and, thus, a higher correlation than would be found if individual-level data were used. Several researchers of school effectiveness and curriculum impact have recently pointed out the deficiencies of using institution-level data for such studies (Aitkin & Longford, 1986; Angoff & Johnson, 1990); these authors specifically point to the fact that institution-level data can overstate the significance of differences among institutions. Social scientists do, however, frequently have an interest in the effects of certain variables upon groups independent of the personal attributes of the individuals comprising the groups; in these cases the analysis of aggregate data is of interest in its own right (Langbein & Lichtman, 1978).

The decision to use institution-level data for this particular study has been made in consideration of one very important fact that suggests that aggregate
COMP data are indeed the data of interest: COMP is intended to provide information about programs rather than individuals. The ACT company has been quite clear in its stance that program scores, that is, means, rather than individual COMP scores, are to be used for interpretation or inference (Steele, 1993). Since the purpose of the test is not to determine how much individual students have learned but rather how well a particular program is impacting groups of students, scores should always be used in aggregate. For this reason, institution-level data were used for this study.

Study Population and Sample

Inferences from the results of research studies involving samples of individual subjects are generally made to a population of individuals. Being a test validity study, however, inferences from the results of this study were made about the test instrument, that is, COMP, rather than a population of users of the test or test-takers. Appropriately, though, one can speak about a population from which subjects for this validity study were chosen. For various reasons, certain restrictions were placed on this population when selecting the group of targeted subjects.

In its largest sense, the population for this study is comprised of colleges and universities that have or continue to administer COMP for purposes of program evaluation. From these institutions, numbering about 500, a particular group of institutions was selected to comprise the actual targeted subjects for this study. Restrictions in choosing the subjects were made in order to control, from the very beginning, gross institutional differences accounted for by two major factors that distinguish institutions of higher learning: (1) institutions are either two-year or four-year, and (2) four-year institutions are either heavily focused on doctoral programs and research (classified by Carnegie as Research Universities I and II and
Doctorate-Granting Universities I and II) or focused on an undergraduate program while supporting a few master's programs (classified by Carnegie as Comprehensive Universities and Colleges I and II and Liberal Arts Colleges I and II). These factors have enormous implications for the institutions' programs, curricula, services, budgets, and campus environments.

The actual group of institutions chosen for this study was four-year colleges and universities that are primarily undergraduate in their mission, that is, not doctoral-granting or research universities. Four-year rather than two-year institutions were chosen because virtually all four-year institutions have general education as part of their undergraduate curriculum, whereas two-year institutions sometimes provide terminal, career-oriented programs that do not include courses in general education; where general education does exist, it is often tailored to enhance the career programs (Gaff, 1983). This fact suggests that the structural elements of general education are quite different for two-year and four-year institutions, and it might be problematic to incorporate both types of institutions into the same study. Four-year institutions that are primarily undergraduate, that is, liberal arts and comprehensive institutions, were chosen because general education in these institutions can be quite different from general education in doctoral-granting and research universities. For example, large doctoral-granting and research institutions make heavier use of graduate teaching assistants than regular faculty in many undergraduate general education courses (Astin, 1993); this would rarely be the case in liberal arts or comprehensive institutions.

Two additional criteria were used to select the sample of institutions. First, institutions were chosen that had used COMP within a three-year time span--academic years 1990-91 through 1992-93. This was done so that data collected on general education programs and institutional characteristics were relatively close to
what test-takers had actually experienced. Using COMP data from several years ago could have meant that a different general education curriculum was operating or even that institutional practices differed and resulted in a changed demographic picture. Secondly, a final scan of testing practices was done at each of the institutions to be sure that the testing samples were reasonably representative of the seniors at the institution. Limited samples, such as test-takers that were all from one major, were not desirable because of a higher probability in these cases that factors related to the limitation could have influenced COMP results at that institution.

Therefore, the targeted sample for this study was 60 institutions that administered the COMP Objective Test to seniors between academic years 1990-91 and 1992-93.

Inherent Design Problems

As noted above, every research study involves certain trade-offs in design and methodology; one's goal is to conduct a study that allows for reasonably valid inferences and fair generalizations to the study population while being cognizant of unavoidable flaws. A brief summary of the flaws of this study seems in order.

First, the operationalization of the general education program was limited to two specific structural components. The rationale has been based on the understanding that these two components best fit the common language that educators have about general education programs. The flaw, however, is that these two components represent only a small portion of the actual curriculum that students experience. COMP scores could in fact be sensitive to the totality of what makes up general education but not necessarily to these two components in isolation. Also, since this study focused on structural components, there is the possibility that the official general education program described in an institution's
catalog or by its academic dean may differ quite substantially from the curriculum that is actually implemented and experienced by its students. Either of these flaws could lead to an error of Type II: falsely concluding that COMP is not a valid measure of program differences.

Second, along with a measure of pre-college achievement, several institutional variables were used in this study to control for things other than the general education program that might be affecting COMP scores. The literature on the effects of college provided the foundation for selecting these particular concomitant variables. The flaw here is that there probably are additional institutional characteristics that affect learning that are not being considered; one rather obvious variable of this type is the quality of teaching in an institution or in a class. Furthermore, the institutional characteristics affecting learning could conceivably vary from institution to institution; there was no way in this study, for example, to control for the one or two colleges that might have extensive internship programs that influence the learning of their students. This flaw could lead to an error of Type I: concluding that COMP scores measure program differences when in fact other confounding factors were overlooked.

Third, the students whose scores were used in this study could have taken more than four years to complete their college degrees. The general education program and institution that some of these students actually experienced may not be the one that is being described by the variables being measured. Institutions and programs have been known to change quite dramatically in a span of just a few years. The flaw, then, is that, once again, all of the explanatory variables, including those depicting the structure of the general education program, may not be accurate descriptors of the students' experiences whose COMP scores were included as the dependent variable. The hope is that the data on the various institutions, while not
100% accurate, approaches an order of magnitude that is still meaningful.

Fourth, the sampling methods used by institutions to collect COMP data used in this study were not all the same. Of the 41 institutions that were subjects for this study, 29 (71%) had required the test of all graduating seniors. Another 7 (17%) had only tested senior students who volunteered to sit for the exam. The remaining 5 (12%) had used some type of sampling plan—either a random sample or one that excluded certain categories of students. The problem is that, while data from multiple institutions are necessary for this type of study, the data are not perfectly comparable across institutions. While one alternative would be to use data only from institutions that, for example, administered the test to the entire population of seniors, the number of institutions would diminish even further and be too small to yield results from which to draw reasonable generalizations. Again, the hope here is that the data from the various institutions, while not collected with 100% consistency, represent the institutions in a way that is still meaningful.

One thing certain is that the students in these samples entered these institutions as first-year students and not as transfer students. This is a known fact for two reasons: (1) those institutions included in this study that had done longitudinal studies administered COMP to the same individuals as seniors that had taken the test when they first entered the college or university as first-year students; and (2) ACT Composite score means, provided for all of the institutions in this study, do not travel from a student's first institution to another in the ACT testing company's data base. Therefore, it is quite certain that there are no transfer students in the testing samples involved in this study.

Data Collection

Dependent variable data, namely COMP scores, were provided by the ACT
company in accordance with their research policies. Each targeted institution was asked to sign a form allowing the ACT company to release their COMP means for this study. (A copy of this form can be found in Appendix A.) Each institution was assured of confidentiality, meaning that institutions would not be mentioned by name in this dissertation. Student confidentiality was not an issue here, because the subject was the institution rather than the individual student. (See Appendix B for confirmation of this important point from the chairperson of the Human Subjects Institutional Review Board at Western Michigan University.) The COMP data were released in the form of institutional means and standard deviations, not as individual scores.

Independent variable data, that is, data about general education programs, were primarily gathered by means of a brief survey that was sent to each institution that agreed to be in the study. (A copy of the survey instrument can be found in Appendix C.) Each institution was asked to specify the number of general education credits/courses required as well as the number of credit/courses required for graduation. In addition, institutions were asked to specify the number of courses that students can choose from to meet the general education requirement. Where responses to particular questions varied within institutions, such as among different fields of study, the values used in the analysis were weighted by the percent of students in these different fields. For example, in one institution, 26% of the students were in baccalaureate programs that required 136 hours to complete and 74% were in programs that required 128 hours to complete; a weighted average was computed and 130 baccalaureate hours was used to represent, for that institution, the number of hours required to graduate.

There is always the possibility that survey items can be misinterpreted or that survey respondents can provide incorrect data. Taking multiple measurements
is good practice as a way to better insure the accuracy and comparability of the data. As illustrated in Table 2, multiple measurements of the general education program characteristics were taken through the use of a survey as well as by examining college catalogs, either hard copy or on microfiche.

Table 2
Sources of Independent and Concomitant Variable Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Survey of Colleges</th>
<th>IPEDS</th>
<th>Peterson's</th>
<th>U.S. News Guide</th>
<th>College Catalogs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of Students Who are Transfers</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Percent of Students Living on Campus</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Number of Students per Faculty Member</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Percent of Faculty With Terminal Degree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Degree of Institutional Control Over General Education Course Selection</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Percent of Total Undergraduate Coursework Devoted to General Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Data on concomitant variables were gathered in several ways; Table 2 shows the various sources of data. The pre-college ACT Composite scores were provided by the ACT company. The remaining institutional variables were gathered from several sources: (a) data from the federal government's Integrated Postsecondary
Education Data System (IPEDS) reported by John Minter, Inc., (b) college 
guidebooks such as Peterson's Four-Year Colleges for 1992, 1993 and 1994, and (c) 
college catalogs both in hard copy and microfiche. Where possible, the data 
themselves as well as data entry were verified by gathering the same data elements 
from multiple sources and on multiple academic years. For example, data on 
number of students per faculty member were gathered for academic years 1990-91, 
1991-92 and 1992-93 as well as from two different sources: (1) Peterson's Four-
Year Colleges, (2) the U.S. News and World Report 1994 College Guide; and (3) 
IPEDS data reported by John Minter, Inc. Where the data differed among academic 
years, an average value was used in the final data analysis. In no case were 
differences from one year to the next so great that an average produced a misleading 
representation of the institution.

Analysis Procedures

Exploratory Data Analysis

The first step in the analysis of the data was to examine frequencies and 
descriptive statistics of all variables involved. In addition, associations among 
variables were examined using scatterplots and correlation coefficients. The 
purpose of such exploratory data analysis is to find preliminary indications of 
trends, patterns, and associations that serve as the foundation for the analyses that 
are to follow (Tukey, 1977).

Multiple Regression Analysis

The second step in the analysis of the data was to regress each of the 
dependent variables in turn on the explanatory variables. This meant that seven
regression models were built, one for the COMP Total means as well as one for each of the subscale means.

The type of regression used was least squares linear regression. Diagnostic procedures were done to confirm the appropriateness of this type of regression for these data. The method of model building used was hierarchical in that explanatory variables were entered cumulatively in a prespecified sequence. The first variable entered was ACT Composite because correlations and prior research indicated that it would account for most of the variance in COMP means. Secondly, the concomitant variables were entered as a group. The third and fourth additions to the model were the variables defining the general education program; the order was determined on the basis of correlations with the dependent variables. The interaction between the two general education variables was also considered for use in the regression.

Hierarchical analysis allows one to determine the addition that each variable or set of variables makes to the $R^2$. Once the models were complete, slopes, $t$-statistics, and partial correlation coefficients were examined to determine the relative importance of each variable and to test each hypothesis. To repeat, the two types of hypotheses were:

1. There is a relationship between an institution’s COMP Total or subscale scores and the amount of control that an institution assumes over students’ selection of their general education courses, measured as the ratio of number of general education courses required to the number of courses from which students may choose in completing their general education requirements.

2. There is a relationship between an institution’s COMP Total or subscale scores and the number of general education courses that students take, measured as the percent of total undergraduate coursework that is devoted to general education.
In all, 14 hypotheses were tested. The criterion used to judge each was that the slope of the independent variable would be statistically different from zero. To determine statistical significance, an alpha level of .10 was used. This is higher than is typically used in social science research, but it was deemed reasonable based on the fact that the sample size was small and the effect size was also expected to be quite small. A higher alpha level essentially requires less stringent evidence to reject a null hypothesis. Small sample sizes and small effect sizes make it harder to find the evidence to reject a null hypothesis even if the null is false. Under such circumstances, a higher, less stringent, alpha level can be used.

Along with examining slopes for statistical significance, the magnitude of the associations between the dependent variables and the independent variables while controlling for the effects of the concomitant variable was examined by computing partial correlation coefficients.

Examination of Outlying Observations

Once the regression models were complete, the next step was to examine outliers and look for any highly influential data points. This is generally done to identify cases for which the overall model does not seem to hold. Such cases can lead to some refinement of the model or can point out other considerations for future research. Outliers and influential data points were identified by computing externally studentized residuals and Cook's distance values.

The next chapter describes the results of the analysis.
CHAPTER IV

RESULTS OF THE ANALYSIS

The intent of the analysis was to test two hypotheses to determine whether or not COMP is sensitive to differences among institutions with respect to two specific characteristics of general education programs. The unit of analysis was the institution, and the sample was 41 liberal arts colleges and comprehensive universities that have used COMP for assessment one or more times between academic years 1990-91 and 1992-93. Analysis of the data involved three steps: (1) exploratory data analysis using histograms, plots, and descriptive statistics, (2) multiple regression analysis, and (3) examination of outlying observations. The purpose of this chapter is to describe the results of the analysis. Conclusions drawn from these results will be presented in Chapter V.

The Final Sample

Of the 60 liberal arts colleges and comprehensive universities targeted for this study, 44 (73.3%) agreed to participate and allow the ACT company to release their COMP scores. Complete data, that is, test scores as well as institutional characteristics and information about general education programs, were collected on all 44 institutions that agreed to participate.

In the end, however, three institutions were eliminated from the final study sample. Two were removed at the recommendation of ACT because their testing samples appeared not be representative of all seniors at these institutions (Steele,
The first institution's testing sample was composed entirely of students from their bachelor in social work program and so was clearly not representative of the entire senior class. The second institution was eliminated because only 8.8% of the students who were invited to sit for the COMP exam actually completed the test, and only half of these were students with ACT Composite scores. This sample was judged to be too different from the entire senior class to be included in the study. In the case of the remaining 42 testing samples, the institutions themselves as well as the director of COMP at the ACT company were reasonably satisfied that the students whose scores were included in the computation of means for this study

Table 3
Attributes of the 41 Institutions Comprising the Final Sample

<table>
<thead>
<tr>
<th>Attribute</th>
<th>No. of Cases</th>
<th>% of Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Control:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private</td>
<td>30</td>
<td>73.2%</td>
</tr>
<tr>
<td>Public</td>
<td>11</td>
<td>26.8%</td>
</tr>
<tr>
<td>Carnegie Classification:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liberal Arts</td>
<td>25</td>
<td>61.0%</td>
</tr>
<tr>
<td>Comprehensive</td>
<td>16</td>
<td>39.0%</td>
</tr>
<tr>
<td>Location:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid-West</td>
<td>20</td>
<td>48.8%</td>
</tr>
<tr>
<td>West</td>
<td>5</td>
<td>12.2%</td>
</tr>
<tr>
<td>South</td>
<td>11</td>
<td>26.8%</td>
</tr>
<tr>
<td>Mid-Atlantic</td>
<td>5</td>
<td>12.2%</td>
</tr>
<tr>
<td>East</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Highest Degree:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baccalaureate</td>
<td>19</td>
<td>46.3%</td>
</tr>
<tr>
<td>Master's</td>
<td>19</td>
<td>46.3%</td>
</tr>
<tr>
<td>Doctorate</td>
<td>3</td>
<td>7.3%</td>
</tr>
<tr>
<td>Total Students:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Under 1,000</td>
<td>18</td>
<td>43.9%</td>
</tr>
<tr>
<td>1,001 to 3,000</td>
<td>14</td>
<td>34.1%</td>
</tr>
<tr>
<td>3,001 to 6,000</td>
<td>4</td>
<td>9.8%</td>
</tr>
<tr>
<td>6,001 to 9,000</td>
<td>3</td>
<td>7.3%</td>
</tr>
<tr>
<td>9,001 to 12,000</td>
<td>2</td>
<td>4.9%</td>
</tr>
</tbody>
</table>
were representative of graduating students at the institution. A third institution was eliminated because no ACT Composite scores were available for any of the students in the COMP testing sample; this institution would necessarily have been eliminated from all of the final regression analyses.

The final sample then was comprised of 41 institutions. Table 3 provides some general information about these colleges and universities.

The 16 institutions that chose not to participate in this study differed from the final sample in several important ways. Names of these institutions were available to the researcher because of publicly available participation lists at the ACT company. Compared to the final sample, a higher percentage of the 16 were public (43.0% compared to 26.8% in the final sample), a higher percentage were comprehensive universities (69.0% compared to 39.0%), and a higher percentage granted degrees above the baccalaureate (75.0% compared to 53.6%). These sixteen institutions also tended to enroll more students, with the median total enrollment of these 16 colleges and universities being roughly 3,800 compared to 1,200 for the final sample. Unfortunately, the composition of the final sample for this study would have been quite different had these additional 16 institutions agreed to participate. The effects of this difference cannot be known.

Exploratory Data Analysis

During the initial stage of the analysis, histograms and descriptive statistics for each of the variables were examined. The purpose was to gain an understanding of the frequency distributions of all variables as well as to identify any dependent variables for which questions about normality might arise. Following this, plots and correlations were examined, looking particularly at the strength and linearity of associations among variables.
Descriptive Statistics of Dependent Variables

None of the histograms of the dependent variables indicated any serious deviations from normality. Descriptive statistics of the COMP Total and six subscale means are shown in Table 4. (Table 4 is showing means of the 41 institutional means that are each based on the institution's own testing sample. Recall from Chapter III that all of the analysis of COMP data was performed on means, not on individual scores.)

One way to judge the representativeness of these testing samples would be to compare these means with national COMP data. Comparing the means of these 41 institutions with a national mean, however, is somewhat difficult because the materials provided by the ACT company about COMP generally report a mean of individual scores rather than a mean of means. Where institutional means are shown, these are typically reported within categories of institutions by ACT Composite score means. A mean of institutional means is reported in one table provided by the ACT company, it being the mean of means from 63 institutions that conducted either two or four-year longitudinal studies using COMP. The senior mean of means reported from this national sample was 183.4 with a standard deviation of 5.8 (American College Testing Program, 1992). This is in fact quite similar to the mean of means in this study of 184.2 with a standard deviation of 4.9. It would appear that, in terms of COMP results, this sample of institutions is somewhat comparable to the larger population of institutions that have used COMP in the recent past.

Descriptive Statistics of Independent Variables

Descriptive statistics for the two independent variables are shown in Table 5.
In addition, bar graphs of these variables are shown in Figures 3 and 4.

Table 4

Descriptive Statistics for Dependent Variables for N = 41 Colleges and Universities

<table>
<thead>
<tr>
<th>Test Scale</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Standard Deviation</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>184.2</td>
<td>184.9</td>
<td>175.4</td>
<td>195.1</td>
<td>4.9</td>
<td>0.4</td>
</tr>
<tr>
<td>The Content Subscales:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functioning Within Social Institutions (FSI)</td>
<td>61.7</td>
<td>61.9</td>
<td>57.4</td>
<td>65.4</td>
<td>1.6</td>
<td>-0.2</td>
</tr>
<tr>
<td>Using Science and Technology (US)</td>
<td>63.4</td>
<td>63.3</td>
<td>60.5</td>
<td>68.8</td>
<td>1.9</td>
<td>0.6</td>
</tr>
<tr>
<td>Using the Arts (UA)</td>
<td>59.4</td>
<td>59.5</td>
<td>55.9</td>
<td>63.4</td>
<td>1.9</td>
<td>0.3</td>
</tr>
<tr>
<td>The Process Subscales:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicating (COM)</td>
<td>52.0</td>
<td>52.1</td>
<td>48.1</td>
<td>55.2</td>
<td>1.8</td>
<td>-0.2</td>
</tr>
<tr>
<td>Solving Problems (SP)</td>
<td>75.4</td>
<td>75.4</td>
<td>72.0</td>
<td>80.0</td>
<td>2.0</td>
<td>0.4</td>
</tr>
<tr>
<td>Clarifying Values (CV)</td>
<td>57.0</td>
<td>56.9</td>
<td>53.0</td>
<td>61.5</td>
<td>1.7</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Recall from Chapter III that the values for the variable degree of institutional control over general education course selection were computed by dividing the number of total general education courses that students are required to complete by
Table 5
Descriptive Statistics for Independent Variables for 
\(N = 41\) Colleges and Universities

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Standard Deviation</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree of Institutional Control Over General Education Course Selection</td>
<td>26.4</td>
<td>21.0</td>
<td>3.0</td>
<td>100.0</td>
<td>21.7</td>
<td>2.1</td>
</tr>
<tr>
<td>Percent of Total Undergraduate Coursework Devoted to General Education</td>
<td>40.6</td>
<td>39.1</td>
<td>26.6</td>
<td>73.4</td>
<td>9.5</td>
<td>1.3</td>
</tr>
</tbody>
</table>

the number of courses from which the institution allows students to select these courses. The resulting ratio was then multiplied by 100 to form a whole number. A value of 100, the highest possible value, indicates that students have no choice in selecting general education courses; in these institutions, every student takes exactly the same courses. The lower the value of this variable, the less control the institution assumes over the courses students take to satisfy their general education requirements. The values for the second variable, percent of total undergraduate coursework devoted to general education, was computed by dividing the number of courses required in general education by the number of total courses required to graduate. For ease of understanding, the resulting ratio was multiplied by 100 and rounded to single decimal precision.

Two institutions in the final sample reported having a core general education program in which, although allowing some exceptions, every student is required to take exactly the same courses to satisfy the general education component of their degree. On the variable measuring the degree of institutional control over the
Figure 3. Bar Graph of Degree of Institutional Control Over General Education Course Selection.

Figure 4. Bar Graph of Percent of Total Coursework Devoted to General Education.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
general education curriculum, these two institutions had a value of 100 resulting from dividing the number of general education courses that students are required to complete by the number of courses the institution allows students to choose from to satisfy the requirement. Again, the resulting ratio was multiplied by 100 and rounded to a single decimal point. The remainder of the institutions in the sample had values on this variable that ranged from a low of 3.0, that is, 667 courses were available to satisfy a 17-course requirement, to a high of 72.4, 29 courses were available to satisfy a 21-course requirement. The mean value of this variable, degree of institutional control over general education course selection, was 26.4. Since the core-like programs were pulling this mean slightly upwards, the value of the median was only 21.0. This indicates that, on average, students in the institutions in this sample are given four to five courses to choose from for every general education course they must complete. Figure 3 displays a bar graph of this variable.

On average, the 41 institutions in the final sample required that 40% of the baccalaureate program be devoted to general education courses. The distribution of this variable was slightly skewed to the right. Of the 41 institutions, 38 required that between 25% and 50% of total coursework be devoted to general education, and three institutions required that more than 56% of the curriculum be devoted to general education. Figure 4 shows a bar graph of this variable. Note the trailing of the three institutions at the upper end of the scale as well as the fact that two categories have no data points.

Descriptive Statistics of Concomitant Variables

Descriptive statistics for the five concomitant variables are displayed in Table 6. For each variable, there was good variability in the distributions and, in
fact, distributions of most variables were nearly normal. There was no variable with extreme outlying values.

Table 6
Descriptive Statistics for Concomitant Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Standard Deviation</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT Composite Mean of Students Who Completed COMP</td>
<td>41</td>
<td>21.9</td>
<td>21.5</td>
<td>18.6</td>
<td>25.7</td>
<td>1.7</td>
<td>0.5</td>
</tr>
<tr>
<td>Percent of Students Who are Transfers</td>
<td>41</td>
<td>15.9</td>
<td>14.0</td>
<td>2.0</td>
<td>47.0</td>
<td>9.9</td>
<td>1.4</td>
</tr>
<tr>
<td>Percent of Students Living on Campus</td>
<td>41</td>
<td>48.8</td>
<td>46.3</td>
<td>6.0</td>
<td>90.3</td>
<td>22.9</td>
<td>0.0</td>
</tr>
<tr>
<td>Number of Students per Faculty Member</td>
<td>41</td>
<td>14.2</td>
<td>12.7</td>
<td>3.4</td>
<td>25.3</td>
<td>5.0</td>
<td>0.4</td>
</tr>
<tr>
<td>Percent of Faculty With Terminal Degree</td>
<td>30</td>
<td>66.0</td>
<td>69.0</td>
<td>36.0</td>
<td>96.0</td>
<td>16.6</td>
<td>0.1</td>
</tr>
</tbody>
</table>

There was one data collection problem concerning the concomitant variables. The variable percent of faculty with terminal degree was only reported in college guidebooks by 30 of the 41 institutions. Exploratory correlations (shown in Table 7) indicated that this variable was not meaningfully associated with any of the COMP scores, and so to avoid shrinking the sample, it was not used in the regression analyses.

Zero-Order Correlations

The next step in the exploratory data analysis was to examine scatterplots.
and zero-order correlations among variables to check for linearity and strength of associations. (See Tables 7 and 8 below.)

Table 7

Zero-Order Correlations Between Dependent Variables and Other Variables

<table>
<thead>
<tr>
<th></th>
<th>COMP</th>
<th>FSI</th>
<th>US</th>
<th>UA</th>
<th>COM</th>
<th>SP</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT Composite</td>
<td>.80***</td>
<td>.71***</td>
<td>.68***</td>
<td>.76***</td>
<td>.76***</td>
<td>.77***</td>
<td>.67***</td>
</tr>
<tr>
<td>Percent of Students Who are Transfers</td>
<td>.17</td>
<td>.04</td>
<td>.27</td>
<td>.09</td>
<td>.02</td>
<td>.14</td>
<td>.31**</td>
</tr>
<tr>
<td>Percent of Students Living on Campus</td>
<td>.14</td>
<td>.14</td>
<td>.09</td>
<td>.15</td>
<td>.17</td>
<td>.14</td>
<td>.31</td>
</tr>
<tr>
<td>Number of Students per Faculty Member</td>
<td>-.08</td>
<td>-.16</td>
<td>.00</td>
<td>-.07</td>
<td>-.16</td>
<td>-.08</td>
<td>-.02</td>
</tr>
<tr>
<td>Percent of Faculty With Terminal Degree</td>
<td>.13</td>
<td>-.03</td>
<td>.20</td>
<td>.10</td>
<td>.07</td>
<td>.10</td>
<td>.17</td>
</tr>
<tr>
<td>Degree of Institutional Control Over General Education Course Selection</td>
<td>-.23</td>
<td>-.31</td>
<td>-.15</td>
<td>-.22</td>
<td>-.35**</td>
<td>-.22</td>
<td>-.09</td>
</tr>
<tr>
<td>Percent of Total Undergraduate Coursework Devoted to General Education</td>
<td>.04</td>
<td>-.04</td>
<td>.07</td>
<td>.05</td>
<td>.02</td>
<td>.07</td>
<td>.04</td>
</tr>
</tbody>
</table>

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

Correlation coefficients showed strong and statistically significant (p < .01) associations between ACT Composite and each of the dependent variables (r = .67 to r = .80). As for correlations between the dependent variables and other variables, small to moderate correlations were found between: (a) the percent of students who are transfers and the Clarifying Values subscale (p < .05); (b) the percent of students...
who are transfers and the Using Science subscale; (c) the percent of students living on campus and the Clarifying Values subscale; and (d) the degree of institutional control over general education course selection and five of the seven scales—COMP Total, Functioning Within Social Institutions, Using the Arts, Communicating (p < .05), and Solving Problems.

Table 8
Zero-Order Correlations Among Concomitant and Independent Variables

<table>
<thead>
<tr>
<th>Variables:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT Composite</td>
<td>(1)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of Students Who are Transfers</td>
<td>(2)</td>
<td>-.11</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of Students Living on Campus</td>
<td>(3)</td>
<td>.30</td>
<td>-.06</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Students per Faculty Member</td>
<td>(4)</td>
<td>-.14</td>
<td>-.10</td>
<td>-.51**</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of Faculty With Terminal Degree</td>
<td>(5)</td>
<td>.21</td>
<td>.06</td>
<td>.32</td>
<td>-.04</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Degree of Institutional Control Over General Education Course Selection</td>
<td>(6)</td>
<td>-.21</td>
<td>.02</td>
<td>.09</td>
<td>.00</td>
<td>-.08</td>
<td>1.00</td>
</tr>
<tr>
<td>Percent of Total Undergraduate Coursework Devoted to General Education</td>
<td>(7)</td>
<td>.10</td>
<td>.04</td>
<td>18</td>
<td>-.41***</td>
<td>-.04</td>
<td>.18</td>
</tr>
</tbody>
</table>

***Significant at the .01 level.

As for associations among the concomitant and independent variables, small to moderate correlations were found between: (a) ACT Composite and the percent
of students living on campus; (b) percent of students living on campus and the number of students per faculty member (p < .01); (c) ACT Composite and the degree of institutional control over general education course selection; and (d) the number of students per faculty member and the percent of total undergraduate coursework devoted to general education (p < .01).

The last series of correlations examined involved the interaction of the two independent variables (degree of institutional control over general education course selection multiplied by percent of coursework devoted to general education.) This interaction term was found to have an extremely large correlation with the first independent variable (r = .97). The high correlation meant that it would add no new information to the regression models, and that, if added, it would cause serious problems associated with multicollinearity. The interaction term was therefore not included in any of the final models.

In addition, numerous plots were examined. Although some showed outlying observations, none showed any indications of nonlinear associations.

Multiple Regression Analysis

Several different types of regression analysis exist along with several different procedures for building regression models. The choice of a particular regression type or procedure should be based on the purposes of the research as well as on the characteristics of the data set. The type of regression used in this study was least squares linear regression. Although the most commonly used form of regression, it may not be appropriate in cases involving nonlinear associations or nonconstant error variance; in such cases, however, data can sometimes be transformed so that the assumptions of least squares linear regression are satisfied. As will be discussed, diagnostic procedures using residual plots confirmed that least
squares linear regression without data transformation was appropriate for these data.

The method of model building used was hierarchical in that explanatory variables were entered cumulatively in a prespecified sequence. Hierarchical analysis allows one to determine the addition that each variable or set of variables makes to the $R^2$. Once the model is complete, slopes, $t$-statistics, and partial correlation coefficients can be used to determine the relative importance of each variable.

The first variable entered was ACT Composite. There are two important reasons for entering this variable first and entering it by itself: (1) correlations and prior research indicate that it would account for most of the variance in COMP means; and (2) the issue has been raised by prior researchers that variables entered after ACT Composite contribute little or nothing to an explanation of COMP scores (Pike, 1989a, 1991). Secondly, the concomitant variables were entered as a group. This was done because the individual contributions of these variables were not important for this study, rather the important thing was to enter them prior to the variables characterizing the general education program. The third addition to the model was the variable defining the degree of institutional control over general education course selection, and the last addition was the variable defining the percent of total coursework devoted to general education. These two variables were entered in the order of their correlation with COMP, degree of institutional control having a stronger correlation with COMP ($r = -.23$) than percent of total coursework devoted to general education ($r = .04$).

This type of hierarchical analysis resulted in all variables entering the model, even those that did not make much of a contribution. This procedure was used because the purpose of the analysis was to build a model that would explain the variance in the dependent variable rather than a model that would be used to predict
Table 9
Hierarchical Analysis Results
of COMP Total Means

<table>
<thead>
<tr>
<th>Step</th>
<th>Variables</th>
<th>$R^2$</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ACT Composite</td>
<td>.63</td>
<td></td>
</tr>
</tbody>
</table>
| 2    | %Transfers  
%On Campus  
Students per Faculty | .70   | .07    |
| 3    | Degree of Control Over General Ed Course Selection                        | .71   | .01    |
| 4    | % of Coursework Devoted to General Education                              | .71   | .00    |

future values of the dependent variable. In the latter situation, it might not be efficient to include and continue to collect data on variables that do not contribute significantly to the model, so the emphasis would be on producing a model with the greatest explanatory power and the least number of variables. In this research, however, the selection of explanatory variables was done prior to the analysis as a result of the review of the literature. Only variables that were known to have some association with educational achievement were chosen for the study. In this way, the effects of the variables defining general education program characteristics could be examined after considering the effects of other variables known to be associated with educational achievement. When it came time for the analysis then, all of the variables needed to be built into the model. Hierarchical analysis accomplished this while insuring that the variables entered the model in some nonrandom manner.

For each dependent variable, a table is provided in the sections that follow showing the impact of the hierarchical analysis upon the $R^2$ values. A second table provides the final regression estimates. The discussion of the results will focus on
the $R^2$ values, slopes, $t$-statistics, and partial correlation coefficients.

Model One: Explaining Differences in COMP Total Means

The first model used COMP Total as the dependent variable. Table 9 shows the cumulative $R^2$ series, and Table 10 shows the resulting regression statistics.

ACT Composite explained 63% of the variance of COMP in this sample ($r = .79; p < .01$). This high degree of association is not surprising; it fully agrees with other research involving COMP and ACT Composite scores. In Chapter II, it was mentioned that correlations as high as .86 have been found between COMP and ACT Composite when analyzing data at the level of the individual. One might in fact have expected the association here to be higher since these data are institutional means rather than individual scores. Correlations of means are inflated due to the fact that the variability of the individual scores around the mean has been eliminated. This can be a problem when the correct unit of analysis is the individual and yet data are only available in aggregate form. For validity research on COMP, however, the institution rather than the individual is the correct unit of analysis. Therefore, in this situation attenuation of correlations is not an issue.

The other concomitant variables accounted for 7% of the explanation of COMP Total. This difference in explanatory power between 63% from ACT Composite and 7% from these three variables indicates that, on the institutional level, average entering student ability by far explains the greatest amount of variance in COMP. As prior research has shown, this was to be expected. The important question is: what explains that part of COMP variance not explained by
Table 10
Regression Estimates of COMP Total Means

<table>
<thead>
<tr>
<th>Variable</th>
<th>Regression Coefficient</th>
<th>Standard Error</th>
<th>t-Value</th>
<th>2-Tailed Prob.</th>
<th>Partial Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT Composite</td>
<td>2.47</td>
<td>.30</td>
<td>8.31</td>
<td>.00</td>
<td>.82</td>
</tr>
<tr>
<td>% Transfers</td>
<td>.13</td>
<td>.05</td>
<td>2.70</td>
<td>.01</td>
<td>.42</td>
</tr>
<tr>
<td>% On-Campus</td>
<td>-.02</td>
<td>.02</td>
<td>-.75</td>
<td>.46</td>
<td>-.13</td>
</tr>
<tr>
<td>Students per Faculty</td>
<td>.01</td>
<td>.12</td>
<td>.06</td>
<td>.96</td>
<td>.01</td>
</tr>
<tr>
<td>Degree of Institutional Control Over General Education Course Selection</td>
<td>-.01</td>
<td>.02</td>
<td>-.42</td>
<td>.68</td>
<td>-.07</td>
</tr>
<tr>
<td>% Coursework Devote to General Education</td>
<td>-.02</td>
<td>.05</td>
<td>-.32</td>
<td>.75</td>
<td>-.06</td>
</tr>
<tr>
<td>Intercept</td>
<td>129.91</td>
<td>7.10</td>
<td>18.29</td>
<td>.000</td>
<td></td>
</tr>
</tbody>
</table>

Total $R^2$: .71  
Adjusted $R^2$: .66

average entering student ability? Are there other variables that account for some of this variance, or must it all be attributed to error? In particular, do any characteristics of general education programs account for any of this variance? These are the critical questions that, when answered, will enhance future discussions of the validity of the COMP instrument. Already it is apparent that, in addition to entering student achievement, three identified institutional characteristics do
contribute a somewhat meaningful amount to an explanation of COMP variance.

The final variables to enter the model were those defining two characteristics of general education. The first, degree of institutional control over general education course selection, added 1% to the explanation of COMP variance. The second variable, percent of total coursework devoted to general education, added nothing. Twenty-nine percent (29%) of the variance of COMP Total was left unexplained by this model.

Once all variables were in the model and before the regression estimates were examined, it was necessary to determine the appropriateness of the model for these data. To do this, various residual plots were examined: (a) a plot of the residuals against the fitted values, (b) plots of the residuals against each of the explanatory variables in the model, and (c) normal probability plots of the residuals. In using residual plots as a diagnostic tool, one is looking for trends or patterns in the scatter of points. In the case of the first two types of plots, a curvilinear pattern would suggest that a linear model may not be appropriate or that transformation of one or more variables might be needed. In normal probability plots, a fairly even diagonal line indicates that the residuals are normally distributed; deviations from this are reason for concern.

The plots of residuals against fitted values and of residuals against explanatory variables already in the model showed no worrisome patterns. There were a couple of outlying points, but this alone is not reason to be concerned about the aptness of the model. The normal probability plot also looked quite acceptable. Samples of these plots are shown in Appendix D.

Table 10 shows the final regression estimates of COMP Total. Two variables were statistically significant (p < .10). For these variables it was possible to reject the null hypothesis that the slope formed by the linear association between
an explanatory variable with the dependent variable is zero. These two variables were: (1) ACT Composite, and (2) the percent of students at the institution who are transfers.

Note that neither of the independent variables, neither the degree of institutional control over general education course selection nor percent of coursework that is devoted to general education, had a statistically significant non-zero slope at the alpha level chosen. Together these two independent variables accounted for virtually none of the variance in COMP Total means.

Testing the significance of slopes and partitioning the total $R^2$ are two ways to examine the relative importance of variables in the model. Yet another is to examine the partial correlation coefficients. These coefficients provide a measure of the association between an explanatory variable and the dependent variable while the effects of other variables in the model are held constant. The partial correlation between COMP Total and the degree of institutional control over general education course selection was -.07 ($p = .34$). Also, the partial correlation between COMP Total and the percent of total coursework devoted to general education was -.06 ($p = .38$). These partial correlations are very small and suggest little or no meaningful association between COMP and these characteristics of general education programs.

**Model Two: Explaining Differences in the Functioning Within Social Institutions Subscale Means**

With this second model, we begin using the subscale means as the dependent variable. Since the subscale scores are summed to compute the COMP Total score, it is important to note that these subscale measures are not independent of the COMP Total. This suggests that characteristics of the first model will in some way be repeated in these subsequent models.
The second model used the Functioning Within Social Institutions (FSI) subscale as the dependent variable. Table 11 shows the cumulative $R^2$ series, and Table 12 provides the resulting regression statistics. Residual and normal probability plots once again indicated that the linear model is appropriate for these data.

Table 11
Hierarchical Analysis Results of Functioning Within Social Institutions (FSI) Means

<table>
<thead>
<tr>
<th>Step</th>
<th>Variables</th>
<th>$R^2$</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ACT Composite</td>
<td>.51</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>% Transfers</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>% On Campus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Students per Faculty</td>
<td>.54</td>
<td>.03</td>
</tr>
<tr>
<td>3</td>
<td>Degree of Control Over General Ed Course Selection</td>
<td>.56</td>
<td>.02</td>
</tr>
<tr>
<td>4</td>
<td>% of Coursework Devoted to General Education</td>
<td>.58</td>
<td>.02</td>
</tr>
</tbody>
</table>

In this model, ACT Composite contributed 51% to the explanation of FSI means, and the other concomitant variables added 3%. Together, the independent variables added 4% to the explanation of FSI. Left unexplained by this model was 42% of the variance.

Only ACT Composite had a slope that was significantly different from zero ($p < .10$). Again, neither of the independent variables had a statistically significant slope.
### Table 12
Regression Estimates of Functioning Within Social Institutions (FSI) Means

<table>
<thead>
<tr>
<th>Variable</th>
<th>Regression Coefficient</th>
<th>Standard Error</th>
<th>t-Value</th>
<th>2-Tailed Prob.</th>
<th>Partial 2-Tailed Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT Composite</td>
<td>.71</td>
<td>.12</td>
<td>5.95</td>
<td>.00</td>
<td>.71</td>
</tr>
<tr>
<td>% Transfers</td>
<td>.02</td>
<td>.02</td>
<td>.92</td>
<td>.37</td>
<td>.16</td>
</tr>
<tr>
<td>% On-Campus</td>
<td>-.01</td>
<td>.01</td>
<td>-.94</td>
<td>.35</td>
<td>-.16</td>
</tr>
<tr>
<td>Students per Faculty</td>
<td>-.06</td>
<td>.05</td>
<td>-1.21</td>
<td>.24</td>
<td>-.20</td>
</tr>
<tr>
<td>Degree of Institutional Control Over General Education Course Selection</td>
<td>-.01</td>
<td>.01</td>
<td>-.97</td>
<td>.34</td>
<td>-.16</td>
</tr>
<tr>
<td>% Coursework Devoted to General Education</td>
<td>-.02</td>
<td>.02</td>
<td>-1.14</td>
<td>.26</td>
<td>-.19</td>
</tr>
<tr>
<td>Intercept</td>
<td>48.25</td>
<td>2.87</td>
<td>16.83</td>
<td>.00</td>
<td></td>
</tr>
</tbody>
</table>

Total $R^2$: .58  
Adjusted $R^2$: .50

Partial correlations showed fairly small associations between the independent variables and the dependent variable (FSI). The partial correlation between FSI and the degree of institutional control over general education course selection was -.16 ($p = .17$), and the partial correlation between FSI and the percent of coursework devoted to general education was -.19 ($p = .13$).
Model Three: Explaining Differences in the Using Science and Technology Subscale Means

The third model used the Using Science and Technology (US) subscale means as the dependent variable. Table 13 shows the cumulative $R^2$ series, and Table 14 shows the resulting regression statistics. Plots once again indicated that the linear model is appropriate for these data.

Table 13
Hierarchical Analysis Results of Using Science and Technology (US) Means

<table>
<thead>
<tr>
<th>Step</th>
<th>Variables</th>
<th>$R^2$</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ACT Composite</td>
<td>.46</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>% Transfers</td>
<td>.60</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>% On Campus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Students per Faculty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Degree of Control Over General Ed Course Selection</td>
<td>.60</td>
<td>.00</td>
</tr>
<tr>
<td>4</td>
<td>% of Coursework Devoted to General Education</td>
<td>.60</td>
<td>.00</td>
</tr>
</tbody>
</table>

In this model, 46% of the variance in US was explained by ACT Composite and 14% by the other concomitant variables. Virtually none of the variance in US was explained by both of the independent variables together. Forty percent (40%) of the variance was left unexplained.

Two variables had slopes that were significantly different from zero ($p < .10$): (1) ACT Composite, and (2) percent of students at the institution who are transfers. Partial correlations between US and the independent variables were
Table 14
Regression Estimates of Using Science and Technology (US) Means

<table>
<thead>
<tr>
<th>Variable</th>
<th>Regression Coefficient</th>
<th>Standard Error</th>
<th>t-Value</th>
<th>2-Tailed Prob.</th>
<th>Partial Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT Composite</td>
<td>.84</td>
<td>.13</td>
<td>6.31</td>
<td>.00</td>
<td>.73</td>
</tr>
<tr>
<td>% Transfers</td>
<td>.07</td>
<td>.02</td>
<td>3.22</td>
<td>.00</td>
<td>.48</td>
</tr>
<tr>
<td>% On-Campus</td>
<td>-.00</td>
<td>.01</td>
<td>-.38</td>
<td>.70</td>
<td>-.06</td>
</tr>
<tr>
<td>Students per Faculty</td>
<td>.05</td>
<td>.05</td>
<td>.99</td>
<td>.33</td>
<td>.17</td>
</tr>
<tr>
<td>Degree of Institutional Control Over General</td>
<td>.00</td>
<td>.01</td>
<td>.05</td>
<td>.96</td>
<td>.01</td>
</tr>
<tr>
<td>Education Course Selection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Coursework Devoted to General Education</td>
<td>.01</td>
<td>.02</td>
<td>.37</td>
<td>.72</td>
<td>.06</td>
</tr>
<tr>
<td>Intercept</td>
<td>43.01</td>
<td>3.19</td>
<td>13.49</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Total $R^2$:</td>
<td>.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$:</td>
<td>.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

extremely small. The partial correlation between US and the degree of institutional control over general education course selection was .01 ($p = .48$), and the partial correlation between US and the percent of coursework devoted to general education was .06 ($p = .36$).
Model Four: Explaining Differences in the Using the Arts Subscale Means

The fourth model had as its dependent variable the Using the Arts (UA) subscale means. Table 15 shows the cumulative $R^2$ series, and Table 16 provides the resulting regression statistics.

Table 15
Hierarchical Analysis Results of Using the Arts (UA) Means

<table>
<thead>
<tr>
<th>Step</th>
<th>Variables</th>
<th>$R^2$</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ACT Composite</td>
<td>.57</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>% Transfers</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>% On Campus Students per Faculty</td>
<td>.61</td>
<td>.04</td>
</tr>
<tr>
<td>3</td>
<td>Degree of Control Over General Ed Course Selection</td>
<td>.61</td>
<td>.00</td>
</tr>
<tr>
<td>4</td>
<td>% of Coursework Devoted to General Education</td>
<td>.61</td>
<td>.00</td>
</tr>
</tbody>
</table>

ACT Composite accounted for 57% of the variance in UA means, and the other concomitant variables added 4%. Again, the independent variables here accounted for virtually none of the variance in UA. Thirty-nine percent (39%) of the variance in UA was unexplained.

In this model, only ACT Composite means had a slope that was significantly different from zero ($p < .10$). Partial correlations between UA and the two independent variables again were extremely small. The correlation between UA and the degree of institutional control was -.06 ($p = .36$), and between UA and the percent of coursework devoted to general education was -.02 ($p = .46$).
Table 16
Regression Estimates of Using the Arts (UA) Means

<table>
<thead>
<tr>
<th>Variable</th>
<th>Regression Coefficient</th>
<th>Standard Error</th>
<th>t-Value</th>
<th>2-Tailed Prob.</th>
<th>Partial Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT Composite</td>
<td>.88</td>
<td>.13</td>
<td>6.74</td>
<td>.00</td>
<td>.76</td>
</tr>
<tr>
<td>% Transfers</td>
<td>.03</td>
<td>.02</td>
<td>1.60</td>
<td>.12</td>
<td>.26</td>
</tr>
<tr>
<td>% On-Campus</td>
<td>-.00</td>
<td>.01</td>
<td>-.44</td>
<td>.66</td>
<td>-.08</td>
</tr>
<tr>
<td>Students per Faculty</td>
<td>.01</td>
<td>.05</td>
<td>.20</td>
<td>.84</td>
<td>.03</td>
</tr>
<tr>
<td>Degree of Institutional Control Over General</td>
<td>-.00</td>
<td>.01</td>
<td>-.37</td>
<td>.71</td>
<td>-.06</td>
</tr>
<tr>
<td>Education Course Selection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Coursework Devoted to General Education</td>
<td>-.00</td>
<td>.02</td>
<td>-.11</td>
<td>.92</td>
<td>-.02</td>
</tr>
<tr>
<td>Intercept</td>
<td>39.77</td>
<td>3.13</td>
<td>12.68</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Total $R^2$:</td>
<td>.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$:</td>
<td>.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model Five: Explaining Differences in the Communicating Subscale Means

With this model, we begin seeing models using the process subscale means as the dependent variables: Communicating, Solving Problems, and Clarifying Values. At the onset, it is worth reminding the reader that these scales are not independent of the content subscales described above: Functioning Within Social Institutions, Using Science and Technology, and Using the Arts. This means that
trends that were seen in the previous three regression models are likely to be repeated in some way in the following three models.

The fifth model had as its dependent variable the Communicating (COM) subscale means. Table 17 shows the cumulative $R^2$ series, and Table 18 displays the resulting regression statistics.

Table 17
Hierarchical Analysis Results of Communicating (COM) Means

<table>
<thead>
<tr>
<th>Step</th>
<th>Variables</th>
<th>$R^2$</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ACT Composite</td>
<td>.58</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>% Transfers &lt;br&gt;% On Campus &lt;br&gt;Students per Faculty</td>
<td>.60</td>
<td>.02</td>
</tr>
<tr>
<td>3</td>
<td>Degree of Control Over General Ed Course Selection</td>
<td>.63</td>
<td>.03</td>
</tr>
<tr>
<td>4</td>
<td>% of Coursework Devoted to General Education</td>
<td>.64</td>
<td>.01</td>
</tr>
</tbody>
</table>

Fifty-eight percent (58%) of the variance of COM means was explained by ACT Composite. The other concomitant variables added 2%, and the independent variables added 4% to the explanation of COM means. Left unexplained was 36% of the variance in COM.

Only ACT Composite had a slope that was significantly different from zero ($p < .10$). The partial correlation between COM and the degree of institutional control over general education course selection was -.26 and significantly different from zero ($p = .07$). The partial correlation between COM and the percent of total coursework devoted to general education was -.09 ($p = .30$).
Table 18

Regression Estimates of Communicating (COM) Means

<table>
<thead>
<tr>
<th>Variable</th>
<th>Regression Coefficient</th>
<th>Standard Error</th>
<th>t-Value</th>
<th>2-Tailed Prob.</th>
<th>Partial Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT Composite</td>
<td>.83</td>
<td>.13</td>
<td>6.60</td>
<td>.00</td>
<td>.75</td>
</tr>
<tr>
<td>% Transfers</td>
<td>.02</td>
<td>.02</td>
<td>.90</td>
<td>.37</td>
<td>.15</td>
</tr>
<tr>
<td>% On-Campus</td>
<td>-.01</td>
<td>.01</td>
<td>-.59</td>
<td>.56</td>
<td>-.10</td>
</tr>
<tr>
<td>Students per Faculty</td>
<td>-.04</td>
<td>.05</td>
<td>-.81</td>
<td>.42</td>
<td>-.14</td>
</tr>
<tr>
<td>Degree of Institutional Control Over General Education Course Selection</td>
<td>-.01</td>
<td>.01</td>
<td>-1.54</td>
<td>.13</td>
<td>-.26</td>
</tr>
<tr>
<td>% Coursework Devoted to General Education</td>
<td>-.01</td>
<td>.02</td>
<td>-.53</td>
<td>.60</td>
<td>-.09</td>
</tr>
<tr>
<td>Intercept</td>
<td>35.33</td>
<td>2.99</td>
<td>11.80</td>
<td>.00</td>
<td></td>
</tr>
</tbody>
</table>

Total $R^2$: .64

Adjusted $R^2$: .57

Model Six: Explaining Differences in the Solving Problems Subscale

Table 19 shows the cumulative $R^2$ series when Solving Problems was the dependent variable. Table 20 shows the resulting regression statistics.

Sixty-percent (60%) of the variance in SP means was explained by ACT Composite. The concomitant variables added 5%, and the independent variables added nothing to the explanation of SP means. Left unexplained was 35% of the
variance in SP.

Table 19
Hierarchical Analysis Results of Solving Problems (SP) Means

<table>
<thead>
<tr>
<th>Step</th>
<th>Variables</th>
<th>$R^2$</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ACT Composite</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>% Transfers % On Campus Students per Faculty</td>
<td>.65</td>
<td>.05</td>
</tr>
<tr>
<td>3</td>
<td>Degree of Control Over General Ed Course Selection</td>
<td>.65</td>
<td>.00</td>
</tr>
<tr>
<td>4</td>
<td>% of Coursework Devoted to General Education</td>
<td>.65</td>
<td>.00</td>
</tr>
</tbody>
</table>

Two variables had slopes that were significantly different from zero ($p < .10$): (1) ACT Composite, and (2) percent of students at the institution who are transfers. Partial correlations between SP and the independent variables were both extremely small and nonsignificant. Between SP and the degree of institutional control over general education course selection the partial correlation was -.07 ($p = .34$), and between SP and the percent of coursework devoted to general education the partial correlation was .01 ($p = .47$).

Model Seven: Explaining Differences in the Clarifying Values Subscale Means

Table 21 shows the cumulative $R^2$ series with Clarifying Values as the dependent variable, and Table 22 displays the resulting regression statistics.

Only 45% of the variance of CV means was explained by ACT Composite, the least amount of any of the models. On the other hand, the other concomitant
variables added the largest amount of any of the models—17%. The independent variables added 1% to the explanation of the CV means. Left unexplained was 37% of the variance.

Table 20
Regression Estimates of Solving Problems (SP) Means

<table>
<thead>
<tr>
<th>Variable</th>
<th>Regression Coefficient</th>
<th>Standard Error</th>
<th>t-Value</th>
<th>2-Tailed Prob.</th>
<th>Partial Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT Composite</td>
<td>.94</td>
<td>.13</td>
<td>7.23</td>
<td>.00</td>
<td>.78</td>
</tr>
<tr>
<td>% Transfers</td>
<td>.04</td>
<td>.02</td>
<td>2.17</td>
<td>.04</td>
<td>.35</td>
</tr>
<tr>
<td>% On-Campus</td>
<td>-.00</td>
<td>.01</td>
<td>-.36</td>
<td>.72</td>
<td>-.06</td>
</tr>
<tr>
<td>Students per Faculty</td>
<td>.02</td>
<td>.05</td>
<td>.31</td>
<td>.76</td>
<td>.05</td>
</tr>
<tr>
<td>Degree of Institutional Control Over General</td>
<td>-.00</td>
<td>.01</td>
<td>-.40</td>
<td>.69</td>
<td>-.07</td>
</tr>
<tr>
<td>Education Course Selection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Coursework Devoted to General Education</td>
<td>.00</td>
<td>.02</td>
<td>.07</td>
<td>.94</td>
<td>.01</td>
</tr>
<tr>
<td>Intercept</td>
<td>54.23</td>
<td>3.10</td>
<td>17.50</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Total $R^2$:</td>
<td>.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$:</td>
<td>.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the final model, two variables had slopes that were statistically significant ($p < .10$): (1) ACT Composite, and (2) percent of students who are transfers. The partial correlation between CV and the degree of institutional control over general
education course selection was .14 (p = .21), and the partial correlation between CV and the percent of coursework devoted to general education was -.04 (p = .41).

Table 21
Hierarchical Analysis Results of Clarifying Values (CV) Means

<table>
<thead>
<tr>
<th>Step</th>
<th>Variables</th>
<th>$R^2$</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ACT Composite</td>
<td>.45</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>% Transfers % On Campus Students per Faculty</td>
<td>.62</td>
<td>.17</td>
</tr>
<tr>
<td>3</td>
<td>Degree of Control Over General Ed Course Selection</td>
<td>.63</td>
<td>.01</td>
</tr>
<tr>
<td>4</td>
<td>% of Coursework Devoted to General Education</td>
<td>.63</td>
<td>.00</td>
</tr>
</tbody>
</table>

Summary of Regression Results

Table 23 summarizes the results of the regression analysis, emphasizing the role in each model of the two independent variables under study.

The first independent variable, degree of institutional control over general education course selection, contributed 3% to the explanation of the variance of the Communicating subscale, 2% to the variance of the Functioning Within Social Institutions subscale, and less than 2% in each of the remaining five regression functions. In no case was its slope significantly different from zero, but its partial correlation with Communicating was significantly less than zero ($r = -.26; p = .07; CI_{90} = -.04, -.45$).
Table 22
Regression Estimates of Clarifying Values (CV) Means

<table>
<thead>
<tr>
<th>Variable</th>
<th>Regression Coefficient</th>
<th>Standard Error</th>
<th>t-Value</th>
<th>2-Tailed Prob.</th>
<th>Partial Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT Composite</td>
<td>.80</td>
<td>.12</td>
<td>6.88</td>
<td>.00</td>
<td>.76</td>
</tr>
<tr>
<td>% Transfers</td>
<td>.07</td>
<td>.02</td>
<td>3.70</td>
<td>.00</td>
<td>.54</td>
</tr>
<tr>
<td>% On-Campus</td>
<td>-.01</td>
<td>.01</td>
<td>-1.19</td>
<td>.24</td>
<td>-.20</td>
</tr>
<tr>
<td>Students per Faculty</td>
<td>.02</td>
<td>.04</td>
<td>.34</td>
<td>.73</td>
<td>.06</td>
</tr>
<tr>
<td>Degree of Institutional Control Over General Education Course Selection</td>
<td>.01</td>
<td>.01</td>
<td>.83</td>
<td>.42</td>
<td>.14</td>
</tr>
<tr>
<td>% Coursework Devoted to General Education</td>
<td>-.00</td>
<td>.02</td>
<td>-.22</td>
<td>.83</td>
<td>-.04</td>
</tr>
<tr>
<td>Intercept</td>
<td>38.81</td>
<td>2.77</td>
<td>13.99</td>
<td>.00</td>
<td></td>
</tr>
</tbody>
</table>

Total $R^2$: .63
Adjusted $R^2$: .57

The second independent variable, percent of total coursework devoted to general education, contributed 2% to the explanation of the Functioning Within Social Institutions subscales, but virtually nothing to the other six regression functions that were developed. In no case was its slope significantly different from zero ($p < .10$), and in no case was its partial correlation with the dependent variable significantly greater or less than zero.
Table 23

Summary of Regression Analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Portion of ( R^2 ) Added</th>
<th>Slope</th>
<th>Partial Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>COMP</td>
<td>Degree of Control</td>
<td>.01</td>
<td>-.01</td>
<td>-.07</td>
</tr>
<tr>
<td>2</td>
<td>FSI</td>
<td>Degree of Control</td>
<td>.02</td>
<td>-.01</td>
<td>-.16</td>
</tr>
<tr>
<td>3</td>
<td>US</td>
<td>Degree of Control</td>
<td>.00</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>4</td>
<td>UA</td>
<td>Degree of Control</td>
<td>.00</td>
<td>-.00</td>
<td>-.06</td>
</tr>
<tr>
<td>5</td>
<td>COM</td>
<td>Degree of Control</td>
<td>.03</td>
<td>-.01</td>
<td>-.26*</td>
</tr>
<tr>
<td>6</td>
<td>SP</td>
<td>Degree of Control</td>
<td>.00</td>
<td>-.00</td>
<td>-.07</td>
</tr>
<tr>
<td>7</td>
<td>CV</td>
<td>Degree of Control</td>
<td>.01</td>
<td>.01</td>
<td>.14</td>
</tr>
<tr>
<td>1</td>
<td>COMP</td>
<td>%Crsework - Gen Ed</td>
<td>.00</td>
<td>-.02</td>
<td>-.06</td>
</tr>
<tr>
<td>2</td>
<td>FSI</td>
<td>%Crsework - Gen Ed</td>
<td>.02</td>
<td>-.02</td>
<td>-.19</td>
</tr>
<tr>
<td>3</td>
<td>US</td>
<td>%Crsework - Gen Ed</td>
<td>.00</td>
<td>.01</td>
<td>.06</td>
</tr>
<tr>
<td>4</td>
<td>UA</td>
<td>%Crsework - Gen Ed</td>
<td>.00</td>
<td>-.00</td>
<td>-.02</td>
</tr>
<tr>
<td>5</td>
<td>COM</td>
<td>%Crsework - Gen Ed</td>
<td>.01</td>
<td>-.01</td>
<td>-.09</td>
</tr>
<tr>
<td>6</td>
<td>SP</td>
<td>%Crsework - Gen Ed</td>
<td>.00</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>7</td>
<td>CV</td>
<td>%Crsework - Gen Ed</td>
<td>.00</td>
<td>-.00</td>
<td>-.04</td>
</tr>
</tbody>
</table>

*Significant at the .10 level (p = .07); CI90 = -.04, -.45.

Degree of Institutional Control Over General Education Course Selection

The first seven hypotheses tested were of the type: There is a relationship between an institution's COMP Total or subscale scores and the amount of control...
that an institution assumes over students' selection of their general education courses, measured as the ratio of number of general education courses required to the number of courses from which students may choose to complete their general education requirements.

The criterion used to judge these hypotheses was that the slope of the independent variable would be significantly different from zero using an alpha level of .10. This alpha level is higher than the typical level used in social science research (.05) but was chosen to achieve more power given that the sample size is small and a small effect size was expected. The null and alternative hypotheses can be expressed as follows:

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

Based on this criterion, none of the first seven null hypotheses could be rejected. In none of the seven models did the degree of institutional control over general education coursework produce a significant non-zero slope. Therefore, one cannot say with certainty that COMP is sensitive to this particular characteristic of general education programs.

**Percent of Coursework Devoted to General Education**

The second set of seven hypotheses tested were: There is a relationship between an institution's COMP Total or subscale scores and the number of general education courses that students take, measured as the percent of total undergraduate coursework that is devoted to general education.

The criterion used to judge these hypotheses was the same as above: the slope of the independent variable would be significantly different from zero ($p < 0.10$).
The null and alternative hypotheses can be expressed as follows:

\[ H_0: \beta_1 = 0 \]
\[ H_a: \beta_1 \neq 0 \]

Based on this criterion, none of the second set of seven models produced a significant non-zero slope. Again, one cannot say with any degree of certainty that COMP is sensitive to this particular characteristic of general education programs.

Examination of Outlying Observations

An examination of outlying observations is good practice in any type of regression analysis, and it seemed particularly important here because of the two or three cases with rather extreme values on the independent variables. (See Figures 3 and 4.) Although these variables did not generally offer anything of significant value to the regression models, there is the unlikely possibility that these outlying values, admittedly unique among general education programs, might have had an inappropriate influence on the results. One would not want to draw conclusions that are largely based on the impact of two or three unusual observations. Determining the influential effects, if any, of these cases on the regression models could only enhance the interpretation of the results, and so seemed an appropriate next step.

To begin, two measures commonly used to diagnose outlying and influential data points were computed: externally studentized residuals and Cook's distance. The first, essentially a standardization of the residuals, detects outlying observations. The second, an overall measure of the combined impact of each case on all of the regression estimates, helps to determine whether any observations are having a major impact on the regression equation such that the regression estimates would be dramatically different if these cases were not present.
The standard criterion for judging outlying observations using externally studentized residuals is to be concerned with any cases whose absolute value exceeds $t(.95; n - p - 1)$. In this case, $t(.95; 33) = 2.03$. (The number of parameters [$p$] is determined by the number of betas that are estimated; in this case, betas were estimated for six explanatory variables and the intercept.) Externally studentized residuals were computed for each of the seven regression models, and six cases were found to exceed the criterion and thus can be considered to be outlying observations. (For the reader who might wish to examine these data, they can be found in Appendix E.)

An additional point should be made about outliers. An outlier is an observation whose actual value on the dependent variable far exceeds its predicted value, either positively or negatively. In this research, the cases mentioned above had actual COMP scores that were much higher or lower than the model predicted they would have. In other words, institutions that were outliers had higher or lower mean COMP scores than would be predicted based on ACT Composite scores, percent of students that are transfers, percent of students living on campus, number of students per faculty member, and two characteristics of general education programs. For these institutions with outlying observations, some other unspecified variable must be influencing general education achievement. Obviously, outliers can be very important because by prodding the researcher to learn something about these observations, one can be directed to other variables to consider in future research.

The standard criterion for judging influential data points using Cook's distance is to be concerned with any cases whose absolute value exceeds $F(.50; p, n - p)$. In this case, $F(.50; 7, 34) = .92$. Cook's distance values were computed for each of the seven regression models (see Appendix F), but none were large enough.
to cause concern. This means that none of the cases were seriously influencing the resulting regression estimates such that the estimates would be different if these cases were eliminated.

All of the diagnostics performed on the regression models indicated that the models developed were appropriate for the data, and no extreme or worrisome cases existed that would necessitate making very guarded conclusions. In the final chapter, conclusions drawn from these results are discussed.

Summary of Results

The results of the analysis can be summarized in the following three points:

1. Of the explanatory variables included in this study, ACT Composite and percent of students at the institution that are transfers were the only ones that had statistically significant slopes in any of the models. ACT Composite, which by far explained the majority of the variance of the dependent variables, produced a significant slope in each of the seven models developed; percent of students that are transfers produced a significant slope in four of the seven models developed.

2. The first independent variable, degree of institutional control over general education course selection, did not produce a significant slope in any of the models; the null hypothesis could not be rejected. It did, however, have a small but significant partial correlation with the Communicating subscale.

3. The second independent variable, percent of coursework devoted to general education, also did not produce a significant slope in any of the models; the second null hypothesis also could not be rejected. This variable did not produce any significant partial correlation coefficients.

The implications stemming from these three results will be discussed more fully in the following chapter.
CHAPTER V

CONCLUSIONS

This study researched the validity of COMP, an instrument used by colleges and universities to assess general education programs. The objective was to ascertain whether or not empirical evidence suggests that COMP is sensitive to differences among such programs. Despite the importance of this issue for COMP test validation, there is general agreement that such sensitivity had not been adequately researched.

The targeted population for this study was 60 liberal arts colleges and comprehensive universities that used COMP between academic years 1990-91 and 1992-93. Forty-one institutions comprised the final sample.

Two characteristics of their general education programs were the major emphasis of this study: (1) degree of institutional control over general education course selection, and (2) percent of total coursework devoted to general education. Four additional variables were used as concomitant variables in the analyses: (1) institutional means of ACT Composite scores taken by students prior to college as a measure of pre-college achievement, (2) percent of students who are transfers, (3) percent of students who live on campus, and (4) number of students per faculty member.

None of the null hypotheses that formed the framework for the analysis were rejected; neither of the general education variables produced statistically significant non-zero slopes in any of the regression models that were developed. Therefore, one cannot say with any degree of certainty that COMP is sensitive to the degree of
control that an institution assumes over general education course selection or to the
relative number of general education courses that students complete. However, a
few small partial correlations were found between the dependent variables and the
general education variables. These are probably worth a bit of attention and
discussion, as are a few other findings that surfaced during the analysis.

Discussion of the Results

The results of the analysis suggest several conclusions as well as comments
and recommendations for future research. First, although neither of the general
education variables produced a significant non-zero slope in any of the regression
models, the degree of institutional control over general education did produce a
fairly small but statistically significant negative partial correlation with the
Communicating subscale ($r = -0.26; p = 0.07; CI_{90} = -0.04, -0.45$). This means that
when the effects of the concomitant variables were removed, namely ACT
Composite scores, percent of students that are transfers, percent of students living
on campus, and number of students per faculty member, this particular general
education characteristic still showed a slight relationship with the Communicating
subscales means. Small but nonsignificant partial correlations were also found
between the degree of institutional control over general education and the
Functioning Within Social Institutions subscale ($r = -0.16$) and between the percent
of coursework devoted to general education and the Functioning Within Social
Institutions subscale ($r = -0.19$).

These partial correlations suggest that, although not large enough to produce
a statistically significant non-zero slope, there is an effect that is at least observable.
It might be well to ask whether it is a plausible finding.

Interestingly, these correlations were negative, suggesting that institutional
COMP means, particularly on the Communicating and Functioning Within Social Institutions subscales, increase as the degree of institutional control over general education course selection decreases. (Due to the matter of nonindependent scales, these conclusions are intended to focus more on the overall COMP instrument than on the individual subscales.) In other words, institutions that give students more freedom in choosing general education courses tend to have higher means, and those institutions that are placing restrictions on the number of courses that can be used to satisfy general education requirements tend to have lower means.

One could propose reasonable explanations for this. For example, one could speculate that institutions that require every student to take exactly the same courses might be enhancing the learning of certain students while actually limiting the development of some of the brighter students causing a kind of "regression towards the mean" effect. This type of program might indeed result in lower scores than those programs in which students take different courses that might push nearly every student to a higher level of development. It is also reasonable to postulate that institutions without restrictions on courses are more often enabling students to develop multiple modes of skills; in the case of communication skills, students might not only be developing their writing skills but also oral, graphic, symbolic and numeric modes of communication. So, the direction of the relationship is plausible as is the possibility that the relationship exists. The important conclusion, therefore, is that these partial correlations, although small, are sufficiently distant from zero to conclude that the findings of this research are not definitive evidence that COMP is not sensitive to program differences. More work needs to be done.

The second conclusion concerns the extent to which COMP correlated with the percent of transfer students in an institution. After taking ACT Composite into account, the slope of the variable defining the percent of transfer students in an
institution was significantly different from zero in the model that regressed COMP Total means as well as in four of the six models that regressed subscale means. As noted in earlier chapters, some researchers have argued that COMP correlates with nothing more than measures of pre-college achievement. Yet, the conclusion from this research is that COMP is sensitive to certain characteristics of institutions, most notably, the percent of students that have transferred from elsewhere.

The slopes produced by the percent of students that were transfers were positive, suggesting that COMP means are higher in institutions that have a higher percentage of students that have transferred from elsewhere. This is particularly interesting because of the fact that the students in these samples were not themselves transfer students. The only reasonable conclusion is that the percent of transfers in an institution has something to do with the educational program offered by the institution, and this program has an effect on all students. If in fact the percent of transfers has an impact on some aspect of the general education program specifically, then this study has indeed found evidence that COMP is sensitive to such differences among general education programs.

Again, is this a plausible finding? Is it possible that as institutions admit more transfers they change their program offerings in ways that affect all students in the institution? Is there an effect occurring from the presence of transfer students that institutions themselves are unaware of? Possibly, institutions with more transfers offer more academic and personal support programs to assist transfers as they adapt to a new college. Perhaps all students then benefit from these types of programs. Perhaps the presence of upper-class transfer students in general education courses improves the quality of the education that all students receive. Perhaps the presence of transfer students, some of whom might be older adult students, lends a maturity to the campus environment that somehow affects students'
learning. This phenomenon is certainly worth further study.

Although these findings regarding the sensitivity of COMP to perhaps one structural aspect of general education and one characteristic of an institution that may in turn affect educational programming are encouraging, they are certainly not conclusive; nor do they lend themselves to simple and straightforward discussion. One of the complexities that must be mentioned concerns the matter of statistical power. The fact of the matter is that due to the small size of the sample and the small correlations between the dependent and the independent variables, statistical power, the ability to detect a significant relationship if one exists, was very low in this study. Even using an alpha level of .10, given a correlation less than .30 and a sample of around 40, the chance of finding a significant relationship if one exists is less than 50%.

Unfortunately, there is no easy way out of this problem. The correlations that were found between the dependent and independent variables were extremely small. In order to have at least a better than 50% chance of detecting a significant relationship in this situation, the sample size would have to increase to several hundred. There are simply not that many institutions of a similar type using COMP as an assessment instrument in any two to three year period of time. The sample could be increased to a few hundred if two-year colleges were included with four-year colleges, but using both types of colleges in one study would cause other serious problems with regard to defining a general education program.

Since the forces of probability are against us, one is left, even more than usual, to rely on some measure of intuitive judgment and common sense. My own conclusion is that the results of this study provide slight evidence that COMP may be sensitive to general education program differences, and additional research of this type would be well worth the effort. This conclusion is based on both the
emergence of small partial correlations between the dependent variables and the
general education variables and the meaningful relationships found between the
dependent variables and the percent of transfer students in an institution. The latter
part of this conclusion assumes that the percent of transfers in an institution affects
educational programming, an assumption that needs further exploration.

The emphasis here must be on the phrase slight evidence. The evidence
found certainly does not call for clear and definitive conclusions. Yet, the
operationalization of general education used in this study was quite restrictive and
statistical power was very low. Further research would benefit from considering
more of the structural components of general education along with characteristics
that are student development oriented. More diverse approaches to general
education have been coming into practice over the last several years, and general
education has come to mean much more than courses required for graduation. A
broader operational definition of general education could in fact produce stronger
and more conclusive results. In addition, it will be important to determine if and
how percent of transfers is affecting educational programming. It will also be
important to investigate other college traits, such as percent of students that are at
the graduate level, percent of students who are of color, or institutional size to see if
and how such traits affect educational programming and, in turn, COMP scores.

Quite recently, three authors (Hurtado, Astin, & Dey, 1991) presented an
empirically based taxonomy of general education programs that came out of
extensive research using college catalogs. They found that colleges and universities
are including many different learning activities within the domain of general
education. Some of these include: (a) orientation activities, (b) the role of faculty in
the academic advising program, (c) integration of the residential and academic
programs, (d) incorporation of women's/gender studies or minority/Third World
studies, (e) internships and cooperative education, (f) community service, and (g) field study, seminars, and independent research. To the extent that these activities enable students to learn to function within social institutions, use science and the arts in their everyday lives, communicate with the written and spoken word, solve problems, and clarify their personal and social values, the COMP instrument would likely be sensitive to differences in such activities among colleges. Indeed, the reason that percent of students that are transfers produced significant non-zero slopes in this research may be because this trait is related to some of the types of learning described above; a most obvious example would be orientation activities. If this is the case, then it explains why, in this study, COMP was found to be sensitive to this institutional characteristic. Clearly, more work needs to be done and more evidence needs to be produced. The chief advantage of this present study is that it has given some indication that COMP may be sensitive to differences among programs, exactly the thing that the test must do to be a valid measure of such programs.

Strengths and Weaknesses of This Study

The greatest strength of this research is that it was the first study in 15 years to do a multi-institutional approach to the validity of COMP as an instrument used to evaluate general education programs. In addition, it was the first ever to attempt to determine the net effects of general education program characteristics upon COMP means. As a result, this study makes a significant contribution to the literature on the validity of COMP and to methodologies for studying the validity of tests designed to assess programs.

There were three important weaknesses to this study. The first is that 16 of the 60 target institutions did not participate. Only four of the 16 actually said they
would not participate for various legitimate reasons; the remaining 12 simply did not respond to several attempts to get a release letter signed. Unfortunately, the final sample would have been quite different had all 60 institutions, or even 56 institutions, agreed to participate. This is a serious weakness, and possibly could be corrected in future research of this type.

The second weakness is that the sample size was very small. In light of the small correlation found between COMP and the independent variables, the sample size undoubtedly contributed to the fact that, with one exception, statistically significant non-zero slopes and partial correlations were not found. The reality of the fact, however, is that with pre-college achievement playing such an important role in post-college achievement, factors such as general education program characteristics are going to produce fairly small partial correlations. With a small effect size, a larger sample is needed to detect statistical significance. Yet, there are simply not that many colleges and universities using COMP, and a sufficiently large sample could never be formed even if the sample included all types of colleges and universities.

The third weakness is that only two fairly narrow components of general education programs were included in this study. This was, however, a realistic starting point that did not prove fruitless. This research will become ever more interesting and useful as additional components of general education are added to the list of independent variables.

Final Remarks

Instruments such as COMP are often one component in an institution's assessment program that might also include majors exams, portfolios, capstone courses, opinion surveys given to students and alumni, and transcript analysis. The
effectiveness or validity of these components as evaluation tools must not be taken for granted; research studies can be designed that can evaluate each of these components.

This research has accomplished two things: (1) it has defined certain methodological principles for studying the validity of tests designed to assess programs rather than individuals; and (2) it has shown that more research needs to be done to validate COMP. The COMP does appear to be sensitive to some differences among institutions, most notably differences in the percent of transfer students that an institution accepts; it remains to be seen if and how this trait affects educational programming. Interesting and promising too is the fact that small associations were found between COMP and one particular characteristic of general education programs: the degree of institutional control over general education course selection. Hopefully, future research will continue these investigations into the validity of COMP using the basic methodological principles that have now been established.
Appendix A

Release Letter Signed by Participating Institutions
August 12, 1993

Dear

I am writing on behalf of Christine Brooks, a doctoral candidate at Western Michigan University. Her doctoral work involves a study of four-year colleges and universities that have participated in the College Outcome Measures Program (COMP) since 1980.

Christine's research requires a comparison of means for samples of seniors with other variables. She would be working with means for institutions rather than with any results for individuals. She has agreed to treat the data as confidential and not report results in a way that would identify any institution with specific results. Based on that understanding, ACT has agreed to support her research by contacting you (and other colleges) for permission to release means for seniors at your institution who took the COMP Objective Test.

If you agree to do so, please sign the permission statement below and return this letter to me. I will provide her with the data she needs. Could you notify me of your decision by or before October 15? There will be no need for you to aggregate samples or generate means yourself. If you wish to discuss her research plans, you can reach her at (207) 725-3797 (Institutional Research Office of Bowdoin College).

Christine will be contacting colleges that agree to participate by telephone in mid-October with brief questions about the structure of their general education programs.

Cordially,

Joe M. Steele, Director
College Outcome Measures Program

Permission is hereby granted for ACT to release to Christine Brooks for doctoral research the group level data on senior samples of students who took the COMP Objective Test at the institution indicated above.

Name: ____________________________
Title: ____________________________
Date: ____________________________

2201 North Dodge Street, P.O. Box 188
Iowa City, Iowa 52243
(319) 337-1000

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Appendix B

Letter From Human Subjects Institutional Review Board
Date: October 8, 1993

To: Christine A. Brooks

From: M. Michele Burnette, Chair

Re: HSIRB Project Number 93-09-07

This letter will serve as confirmation that your research project entitled "An examination of the sensitivity of the COMP (College Outcomes Measures Program) instrument to differences among general education programs in liberal arts colleges and comprehensive universities" has been reviewed under the exempt category by the Human Subjects Institutional Review Board.

I concur that the research does not fall under the HSIRB policy since you will be collecting institutional data only; hence, this project does not require HSIRB approval.

Should you have any concerns/questions, please contact Michele L. Rosa at the HSIRB office (387-8293).
Appendix C

Structure of General Education Questionnaire
STRUCTURE OF GENERAL EDUCATION QUESTIONNAIRE

For purposes of this questionnaire, general education is considered to be that part of the undergraduate curriculum that is required of students.

| Name of Institution: |  |
| Name of Person Completing This Survey: |  |
| Title: |  |
| Address: |  |
| Phone: | FAX: |

What amount of the entire undergraduate curriculum is devoted to general education?

Indicate the number of credit hours and/or courses required in your general education program and the total number of hours and/or courses required for graduation. If these numbers differ for different areas of study (e.g., engineering, business, etc.), please first indicate the numbers that hold for the majority of students at your institution. Then, provide these numbers for any curricular areas for which they may differ. Please attach extra sheets if necessary.

<table>
<thead>
<tr>
<th></th>
<th>Hours</th>
<th>Courses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Required in general education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Required to graduate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This differs for the following area of study: __________________

What percent of undergraduate students are enrolled in this area? ____% |

<table>
<thead>
<tr>
<th></th>
<th>Hours</th>
<th>Courses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Required in general education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Required to graduate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This differs for the following area of study: __________________

What percent of undergraduate students are enrolled in this area? ____% |

<table>
<thead>
<tr>
<th></th>
<th>Hours</th>
<th>Courses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Required in general education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Required to graduate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Are the above hours?  semester hours _____  quarter hours _____

How is the general education requirement at your institution structured?

If your general education program is essentially a core program, that is, one that requires all students to take the same set of courses, please check here: _____

If your general education program is essentially a distribution program, that is, one that allows students to choose the courses that will satisfy certain areas of general education requirements, please answer the following:

How many courses do the majority of students have to choose from to complete their general education requirements? _______

This differs for the following area of study: ____________
What percent of undergraduate students are enrolled in this area? ____%
How many courses do these students have to choose from to complete their general education requirements? _______

**************

Please return to me by FAX: (207) 725-3024
Or return by U.S. Mail to: Office of Institutional Research
Bowdoin College
Brunswick, ME 04011
If you have questions, please call me at: (207) 725-3797
Appendix D

Sample Residual and Probability Plots
Sample Plot One: Plot of Residuals Against Predicted Values
Sample Plot Two: Normal Probability Plot of Residuals
Sample Plot Three: Plot of Residuals Against ACT Composite Means
Sample Plot Four: Plot of Residuals Against Degree of Institutional Control Over General Education
Sample Plot Five: Plot of Residuals Against Percent of Total Coursework That is General Education
Appendix E
Externally Studentized Residuals
### Externally Studentized Residuals for COMP Total and Subscale Means

<table>
<thead>
<tr>
<th>Case Number</th>
<th>COMP Total</th>
<th>FSI</th>
<th>US</th>
<th>UA</th>
<th>COM</th>
<th>SP</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.70</td>
<td>0.31</td>
<td>2.19</td>
<td>1.25</td>
<td>0.80</td>
<td>1.28</td>
<td>1.79</td>
</tr>
<tr>
<td>2</td>
<td>-1.03</td>
<td>-0.80</td>
<td>-0.16</td>
<td>-1.33</td>
<td>-0.17</td>
<td>-0.91</td>
<td>-1.27</td>
</tr>
<tr>
<td>3</td>
<td>1.34</td>
<td>3.65</td>
<td>0.32</td>
<td>0.18</td>
<td>0.65</td>
<td>2.37</td>
<td>0.28</td>
</tr>
<tr>
<td>4</td>
<td>-0.10</td>
<td>-0.96</td>
<td>0.17</td>
<td>0.10</td>
<td>-0.84</td>
<td>0.21</td>
<td>0.37</td>
</tr>
<tr>
<td>5</td>
<td>0.12</td>
<td>0.28</td>
<td>-0.01</td>
<td>0.10</td>
<td>-0.33</td>
<td>0.68</td>
<td>0.32</td>
</tr>
<tr>
<td>6</td>
<td>1.31</td>
<td>0.99</td>
<td>1.34</td>
<td>0.69</td>
<td>1.66</td>
<td>0.99</td>
<td>0.64</td>
</tr>
<tr>
<td>7</td>
<td>0.74</td>
<td>1.16</td>
<td>-0.33</td>
<td>1.32</td>
<td>0.96</td>
<td>0.51</td>
<td>-0.08</td>
</tr>
<tr>
<td>8</td>
<td>-1.14</td>
<td>-0.86</td>
<td>-0.37</td>
<td>-1.26</td>
<td>0.02</td>
<td>-0.95</td>
<td>-1.77</td>
</tr>
<tr>
<td>9</td>
<td>-0.63</td>
<td>-0.64</td>
<td>0.36</td>
<td>-1.05</td>
<td>-0.57</td>
<td>-0.73</td>
<td>-0.01</td>
</tr>
<tr>
<td>10</td>
<td>-0.10</td>
<td>0.01</td>
<td>-1.15</td>
<td>0.59</td>
<td>-1.04</td>
<td>0.01</td>
<td>0.78</td>
</tr>
<tr>
<td>11</td>
<td>-0.83</td>
<td>-1.69</td>
<td>0.43</td>
<td>-1.10</td>
<td>-1.14</td>
<td>-0.65</td>
<td>-0.03</td>
</tr>
<tr>
<td>12</td>
<td>0.71</td>
<td>1.00</td>
<td>2.07</td>
<td>-0.82</td>
<td>-0.04</td>
<td>0.45</td>
<td>1.48</td>
</tr>
<tr>
<td>13</td>
<td>-1.37</td>
<td>-1.28</td>
<td>-1.36</td>
<td>-0.66</td>
<td>-2.45</td>
<td>-1.13</td>
<td>-0.04</td>
</tr>
<tr>
<td>14</td>
<td>-1.08</td>
<td>-0.79</td>
<td>-0.96</td>
<td>-0.73</td>
<td>-0.45</td>
<td>-1.26</td>
<td>-1.32</td>
</tr>
<tr>
<td>15</td>
<td>0.13</td>
<td>-0.35</td>
<td>0.33</td>
<td>0.21</td>
<td>0.55</td>
<td>-0.56</td>
<td>0.28</td>
</tr>
<tr>
<td>16</td>
<td>-0.51</td>
<td>0.28</td>
<td>-1.14</td>
<td>0.07</td>
<td>0.35</td>
<td>-0.39</td>
<td>-1.39</td>
</tr>
<tr>
<td>17</td>
<td>-0.07</td>
<td>0.50</td>
<td>-1.43</td>
<td>1.15</td>
<td>-0.32</td>
<td>0.13</td>
<td>-0.30</td>
</tr>
<tr>
<td>18</td>
<td>0.12</td>
<td>0.38</td>
<td>0.09</td>
<td>-0.20</td>
<td>0.26</td>
<td>0.25</td>
<td>-0.33</td>
</tr>
<tr>
<td>19</td>
<td>-0.61</td>
<td>0.56</td>
<td>-1.20</td>
<td>-0.29</td>
<td>-0.15</td>
<td>-0.78</td>
<td>-0.54</td>
</tr>
</tbody>
</table>
### Externally Studentized Residuals for COMP Total and Subscale Means--continued

<table>
<thead>
<tr>
<th>Case Number</th>
<th>COMP Total</th>
<th>FSI</th>
<th>US</th>
<th>UA</th>
<th>COM</th>
<th>SP</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>-1.38</td>
<td>-1.17</td>
<td>-1.32</td>
<td>-0.46</td>
<td>-0.95</td>
<td>-1.79</td>
<td>-1.13</td>
</tr>
<tr>
<td>21</td>
<td>0.12</td>
<td>-0.52</td>
<td>0.06</td>
<td>0.30</td>
<td>-0.39</td>
<td>0.60</td>
<td>0.15</td>
</tr>
<tr>
<td>22</td>
<td>1.37</td>
<td>1.27</td>
<td>1.34</td>
<td>0.68</td>
<td>1.18</td>
<td>1.40</td>
<td>0.82</td>
</tr>
<tr>
<td>23</td>
<td>0.09</td>
<td>-0.49</td>
<td>0.30</td>
<td>0.13</td>
<td>-0.11</td>
<td>0.24</td>
<td>0.56</td>
</tr>
<tr>
<td>24</td>
<td>0.65</td>
<td>0.22</td>
<td>0.15</td>
<td>1.04</td>
<td>1.60</td>
<td>0.13</td>
<td>0.17</td>
</tr>
<tr>
<td>25</td>
<td>-1.55</td>
<td>-2.37</td>
<td>-0.57</td>
<td>-0.88</td>
<td>-1.02</td>
<td>-1.45</td>
<td>-1.22</td>
</tr>
<tr>
<td>26</td>
<td>0.25</td>
<td>0.05</td>
<td>0.28</td>
<td>0.13</td>
<td>-0.30</td>
<td>0.20</td>
<td>0.41</td>
</tr>
<tr>
<td>27</td>
<td>1.18</td>
<td>0.41</td>
<td>0.65</td>
<td>1.05</td>
<td>0.08</td>
<td>1.52</td>
<td>1.98</td>
</tr>
<tr>
<td>28</td>
<td>-0.65</td>
<td>-1.51</td>
<td>0.52</td>
<td>-0.71</td>
<td>-1.01</td>
<td>-0.79</td>
<td>0.06</td>
</tr>
<tr>
<td>29</td>
<td>-1.70</td>
<td>-1.38</td>
<td>-1.59</td>
<td>-1.10</td>
<td>-1.56</td>
<td>-1.34</td>
<td>-1.37</td>
</tr>
<tr>
<td>30</td>
<td>0.21</td>
<td>-0.72</td>
<td>-0.44</td>
<td>0.76</td>
<td>0.00</td>
<td>-0.69</td>
<td>1.10</td>
</tr>
<tr>
<td>31</td>
<td>2.26</td>
<td>0.94</td>
<td>2.71</td>
<td>1.32</td>
<td>2.44</td>
<td>1.47</td>
<td>2.27</td>
</tr>
<tr>
<td>32</td>
<td>2.19</td>
<td>1.90</td>
<td>1.51</td>
<td>1.69</td>
<td>1.98</td>
<td>2.25</td>
<td>1.12</td>
</tr>
<tr>
<td>33</td>
<td>-0.63</td>
<td>-0.81</td>
<td>0.91</td>
<td>-1.68</td>
<td>-2.20</td>
<td>-0.45</td>
<td>1.11</td>
</tr>
<tr>
<td>34</td>
<td>1.89</td>
<td>1.13</td>
<td>0.22</td>
<td>3.93</td>
<td>1.67</td>
<td>1.78</td>
<td>0.54</td>
</tr>
<tr>
<td>35</td>
<td>-0.48</td>
<td>0.94</td>
<td>-0.93</td>
<td>-0.71</td>
<td>0.51</td>
<td>-0.46</td>
<td>-1.25</td>
</tr>
<tr>
<td>36</td>
<td>0.59</td>
<td>0.66</td>
<td>0.57</td>
<td>0.02</td>
<td>0.64</td>
<td>0.57</td>
<td>0.29</td>
</tr>
<tr>
<td>37</td>
<td>0.38</td>
<td>0.03</td>
<td>-0.11</td>
<td>0.70</td>
<td>0.22</td>
<td>0.88</td>
<td>0.11</td>
</tr>
<tr>
<td>38</td>
<td>-1.00</td>
<td>0.24</td>
<td>-1.36</td>
<td>-1.00</td>
<td>-0.19</td>
<td>-0.91</td>
<td>-1.57</td>
</tr>
<tr>
<td>39</td>
<td>-0.99</td>
<td>-0.37</td>
<td>-0.48</td>
<td>-1.37</td>
<td>-0.12</td>
<td>-1.29</td>
<td>-1.53</td>
</tr>
<tr>
<td>40</td>
<td>-0.36</td>
<td>-0.26</td>
<td>-0.09</td>
<td>-0.58</td>
<td>0.06</td>
<td>-0.39</td>
<td>-0.38</td>
</tr>
<tr>
<td>41</td>
<td>-0.85</td>
<td>0.32</td>
<td>-1.25</td>
<td>-0.63</td>
<td>-0.17</td>
<td>-0.61</td>
<td>-1.30</td>
</tr>
</tbody>
</table>
Appendix F

Cook's Distance Values
<table>
<thead>
<tr>
<th>Case Number</th>
<th>COMP Total</th>
<th>FSI</th>
<th>US</th>
<th>COM</th>
<th>SP</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.26</td>
<td>0.01</td>
<td>0.41</td>
<td>0.00</td>
<td>0.06</td>
<td>0.29</td>
</tr>
<tr>
<td>2</td>
<td>0.04</td>
<td>0.03</td>
<td>0.00</td>
<td>0.07</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>3</td>
<td>0.08</td>
<td>0.46</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.15</td>
</tr>
<tr>
<td>4</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>0.03</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>11</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>12</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>13</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>14</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>15</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>16</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>17</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>18</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>19</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Cook's Distance Values for COMP Total and Subscale Means

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Case Number</th>
<th>COMP Total</th>
<th>FSI</th>
<th>US</th>
<th>UA</th>
<th>COM</th>
<th>SP</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.09</td>
<td>0.06</td>
<td>0.08</td>
<td>0.01</td>
<td>0.04</td>
<td>0.14</td>
<td>0.06</td>
</tr>
<tr>
<td>21</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>22</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>23</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>24</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
<td>0.16</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>25</td>
<td>0.27</td>
<td>0.58</td>
<td>0.04</td>
<td>0.09</td>
<td>0.12</td>
<td>0.24</td>
<td>0.17</td>
</tr>
<tr>
<td>26</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>27</td>
<td>0.05</td>
<td>0.01</td>
<td>0.02</td>
<td>0.04</td>
<td>0.00</td>
<td>0.08</td>
<td>0.13</td>
</tr>
<tr>
<td>28</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>29</td>
<td>0.13</td>
<td>0.09</td>
<td>0.11</td>
<td>0.06</td>
<td>0.11</td>
<td>0.08</td>
<td>0.09</td>
</tr>
<tr>
<td>30</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>31</td>
<td>0.04</td>
<td>0.01</td>
<td>0.06</td>
<td>0.02</td>
<td>0.05</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>32</td>
<td>0.11</td>
<td>0.08</td>
<td>0.05</td>
<td>0.07</td>
<td>0.09</td>
<td>0.11</td>
<td>0.03</td>
</tr>
<tr>
<td>33</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.09</td>
<td>0.14</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>34</td>
<td>0.08</td>
<td>0.03</td>
<td>0.00</td>
<td>0.27</td>
<td>0.07</td>
<td>0.07</td>
<td>0.01</td>
</tr>
<tr>
<td>35</td>
<td>0.01</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>36</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>37</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>38</td>
<td>0.02</td>
<td>0.00</td>
<td>0.03</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>39</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>40</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>41</td>
<td>0.01</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


Schomberg, S. F., Hendel, D. D., & Bassett, C. L. (1982). *Testing the Usefulness of the college outcomes measures project (COMP) to distinguish among seniors in university baccalaureate programs*. Results of field study support by the American College Testing Program. Manuscript provided by authors.


