ED 338 141   HE 024 982

AUTHOR	Ratcliff, James L.; And Others
TITLE	Development and Testing of the Cluster-Analytic Model to Identify Coursework Patterns Associated with General Learned Abilities of College Students: Report on Evergreen State College Combined Samples.
INSTITUTION	Pennsylvania State Univ., University Park. Center for the Study of Higher Education.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC. Office of Research.
PUB DATE	Jul 91
CONTRACT	OERI-R-86-0016
NOTE	162p.; For a related document, see HE 024 981.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Cluster Analysis; *College Outcomes Assessment; College Seniors; *Curriculum Research; *Educational Assessment; General Education; Higher Education; *Learning Processes; *Outcomes of Education; *Research Methodology; Scores; Statistical Analysis
IDENTIFIERS	Evergreen State College WA

ABSTRACT

        This report presents the findings of the application
of the Differential Coursework Patterns (DCP) cluster analytic model
to the tests and transcripts of a combined sample of 1987-88 and
1988-89 graduating seniors at Evergreen State College (Washington).
The project found that enrollment in different patterns of programs
was consistently associated with gains in the general learned
abilities of undergraduate students. Other findings indicated that
the development of general learned abilities did not have an exact
one-to-one relationship with departmental categories; that the
development of general learned abilities was not confined to the
lower division; and that there was little formal monitoring and
description of the curriculum in terms of general learned abilities.
The report is divided into seven sections, each describing one of the
following research components: the theoretical basis for the
differential coursework hypothesis including the development of the
cluster analytic model as a means of testing the hypothesis; the
procedures and methodology used; the major characteristics of the two
samples of graduating seniors whose transcripts and test scores
provided the data analyzed; characteristics of the programs found on
the student transcripts in terms of the criterion variables; the
intended curriculum of the institution, its goals, curriculum
organization, and structure; the findings; and a summary,
conclusions, and recommendations. Contains 199 references.
(Author)

PENNSTATE

ED338141

Development & Testing of the Cluster-Analytic Model to Identify Coursework Patterns Associated with General Learned Abilities of College Students: Evergreen State College Report on Combined Samples

July 1991

Contract No. OERI-R-86-0016
Project No. R-117G10037
U.S. Department of Education
Office of Educational Research & Improvement

James L. Ratcliff

with

Steven Hoffman
Elizabeth A. Jones

Center for the Study of Higher Education
The Pennsylvania State University
403 S. Allen Street, Suite 104
University Park, PA 16801-5202

An Equal Opportunity University

HE 024 982

2 **BEST COPY AVAILABLE**

Development and Testing of a Cluster-Analytic Model

for Identifying Coursework Patterns

Associated with General Learned Abilities of College Students:

Report on Evergreen State College Combined Samples

July 1991

James L. Ratcliff

with

Steven Hoffman
Elizabeth Jones

Center for the Study of Higher Education
The Pennsylvania State University
403 South Allen Street, Suite 104
University Park, PA 16801-5202

TABLE OF CONTENTS

iii

6

THE DIFFERENTIAL COURSEWORK PATTERNS PROJECT

AND THE DEVELOPMENT OF THE CLUSTER ANALYTIC MODEL

The purpose of the "Differential Coursework Patterns (DCP) Project" was to determine the effect of different patterns of college coursework on the general learned abilities of students. To accomplish this end, a model was developed for linking what coursework students took in college with what they learned in college . The result was the Cluster Analytic Model. The Model groups courses appearing on student transcripts according to the distribution of assessment scores of those students. The Model uses precollege indicators of learning to control for incoming student ability. It uses transcripts, rather than formal course or degree requirements as the representation of the college curriculum. The Model can use any number of assessment measures, including both quantitative and qualitative data.

Over the past five years, the DCP Project team has de;eloped and tested the Cluster Analytic Model in a variety of college and university settings. Consistently, the Model has identified and correctly classified approximately 8 of every 10 courses undergraduates took according to the multiple measures of student learning used. To date, the Model has employed such criterion measures as the 9 item-types of the GRE General Test, Kolb's Learning Styles Inventory, a locally-developed instrument measuring the students' perception of the difficulty of the courses, and the ACT COMP examination subscores. Discriminant analyses of the course groupings according to these measures indicated an average correct classification rate of over 80 percent.

Summary of the Cluster Analytic Model. The basic steps in the DCP cluster analytic model are summarized as follows. For simplicity of example, SAT scores are used as measures of incoming student ability. GRE item-type scores are used as measures of general learning. Coursework patterns are derived from the transcripts of graduating seniors at each of six postsecondary institutions First, student achievement is determined by removing the predicted effect of SAT sub-scores on GRE item-type scores; the 9 item-type residuals serves as 9 criterion variables of general learned ability during the student's baccalaureate years. Second, the mean residual scores of students enrolling in each course is determined and attributed to that course. Third, courses are taxonomized (or grouped into patterns) using one of two cluster analysis procedures according to the 9 GRE residual scores of the students. Fourth, the secondary validity of the clusters and the contribution of each of the 9 item-type residuals to the clusters are established by discriminant analysis. Fifth, the resulting patterns are described in terms of the combinations and sequences in which the students enrolled in them and the predominant curriculum characteristics of the courses in the resulting pattern. Lastly, the coursework patterns are examined to determine the extent to which they are associated with a discipline or set of disciplines and the general education requirements of the college or university.

Following these procedures produces a set of hypothesized relationships between coursework patterns and measures of general learned abilities. To provide primary validation of these relationships, a second sample of graduating seniors has been drawn from each participating institution, and the procedures

of the cluster analytic model were repeated to determine the extent to which the coursework patterns and their relationship with the criterion measures of general learned ability were replicated. In one participating institution, repeated samples have been drawn to determine enrollment trends and their effect over time.

The conceptual framework for the cluster analytic model was derived from a review of selected literature. This review revealed that no one curricular model and no one analytical process clearly identified the effect of differential coursework on general learned abilities. Development and testing of an analytic model for studying the differential effect of coursework in a complex curricular organization was required. A conceptual-empirical approach was adopted for the examination of the research problem, wherein a conceptual framework of student course decisions and selections guided the empirical search for coursework patterns associated with gains in general learned abilities.

Findings of the Research. The following is an excerpt of the findings and conclusions from the combined first and second samples of graduating seniors at one of the institutions participating in the DCP Project, Evergreen State College. The two college student samples were similar to the population of graduating seniors for each year in terms of SAT scores, majors and demographic characteristics. The two samples were also similar to one another. There was no significant year-to-year variation in student characteristics, nor was there major variation in the characteristics of the samples relative to the population. The samples were small and were combined to permit further analysis.

The growth in learning displayed in the test score results varied between the prior two Evergreen samples. Generally, Evergreen students showed more gains in learning in Data Interpretation, Quantitative Comparisons, and Reading Comprehension. In the analysis of courses taken by 2 or more students, these three item-types explained large proportions of the score variance. In each of these analyses, these GRE item-types proved to explain most of the gains in student learning. Taking different coursework produces different effects in general learned abilities, and those effects varied among the students of the 1987-88 and 1988-89 classes of graduating seniors.

What does the Cluster Analytic Model tell us about assessment? The Cluster Analytic Model uses multiple measures of assessment. It provides colleges with information regarding the extent of variation in student assessment results that is explained by any one of the measures used. This information can be helpful in a number of ways. Faculty and administrators need not decide on an ideal set of assessment measures. The extent to which such measures may overlap in describing student learning can be identified. The mix of assessment measures appropriate to the goals of the college and the characteristics of the student population can be continuously monitored. When students show small amounts of growth on an indicator of student learning, either the college can develop strategies for improving student learning in the area identified, or discard the measure as inappropriate to the college and its students. The Cluster Analytic Model provides useful information to the college about the mix of assessment measures that reflects what the students learn and what the college intends to teach them.

8

What does the Cluster Analytic Model tell us about the curriculum? The
Cluster Analytic Model is a tool ideally suited to institutions of higher
education with a distributional general education requirement and a wide array
of programs, electives and majors. For example, if one of the assessment
measures a college selects is a test of analytic reasoning, then the Cluster
Analytic Model can identify those groups of courses that students took who
showed significant improvement in that area of general learning. Furthermore,
the student population can be subdivided into high ability and low ability
students, by gender, race or ethnicity, or by major. Then the Model can
identify if the coursework associated with gains in learning among the total
group is the same as that for the subgroups. Such appropriate information is
valuable to curriculum planners. Courses in the general education sequence not
found to be associated with gains in student learning can be revised, enhanced
or dropped. Courses outside the general education requirements that contribute
to gains in student learning can become candidates for inclusion in the general
education curriculum. The extent to which general education courses affect the
learning of both high ability and low ability students has relevance in deciding
how wide ranging the distributional options should be or whether a core
curriculum is appropriate for the students and the educational goals of the
institution.

Such actions presuppose that the measures used reflect the purposes of
general education at the institution. In the DCP Project, none of the
institutions had agreed upon measures to assess general education.
Consequently, the General Test of the Graduate Record Examination was used.
This examination measures cognitive development primarily, rather than specific
content learning.

How can the Cluster Analytic Model help with advising students? By linking
the coursework students take with their improvement in learning, the Cluster
Analytic Model can be particularly valuable in advising students. First, it
takes advising beyond the mere listing of formal degree requirements to the
identification of those specific courses in which students of comparable
interests, abilities and achievement have enrolled. Given several years of
assessment data linked to the transcripts of graduating seniors, the Model can
identify an array of courses taken by students who showed the largest gains in
general learning in college. The Model is amenable to the development of a
microcomputer-based advising system utilizing a relational database of prior
students coursetaking patterns and assessment results. Such a computer-based
advising system would yield an array of effective coursework tailored to the
abilities and interests of individual students and within the parameters of
institutional degree requirements. Such a computer-based advising system has
been proposed as a logical extension of the DCP Project research.

# INTRODUCTION

The Differential Coursework Patterns Project

and the Development of a Cluster Analytic Model

The purpose of the "Differential Coursework Patterns (DCP) Project" is to determine the effects associated with different patterns of college coursework on the general learned abilities of students. Student samples were drawn as volunteer graduating seniors from 5 geographically and curricularly diverse colleges and universities. Subsequent analysis showed that the samples approximated the distribution of Scholastic Aptitude Test (SAT) scores, majors, and other socio-economic characteristics of the population of graduating seniors at each of the institutions. The precollege general learned abilities of each sample of students was controlled in the research. The effect of SAT scc es on the postcollege measures of student learning was partialled from the ana ysis; the remaining scores were used as criterion variables of student achievement. Thus, exiting student general learned abilities were operationally defined by the residual differences from the predicted and observed scores on 9 types of learning ("item-types") measured by the General Test of the Graduate Record Examination (GRE).

The GRE General Test consists of three sections: verbal, quantitative and analytical; within each section are specific types of test items (i.e. Verbal: analogy items; Quantitative: quantitative comparisons; Analytic: logical reasoning). There is a total of 9 item-types within the General Test, and the 9 residual differences from the predicted and observed scores of students consti- tuted the primary measures of exiting student achievement in general learning. Thus defined, exiting student achievement served as a metric in the analysis of

10

the differential coursework represented in the transcripts of each institutional sample.

Six institutions of higher education participated in the DCP project. Each institution represented a different Carnegie classification. The institutions were: Clayton State College (Morrow, GA: a public community college at the time the students in the sample were enrolled); Evergreen State College (Olympia, WA: public liberal arts college II); Georgia State University (Atlanta, GA: doctoral-granting university I); Ithaca College (Ithaca, NY: comprehensive college I); Mills College (Oakland, CA: private liberal arts college I); and Stanford University (Stanford, CA: private research university I).

The approach taken in the DCP project was empirical. No a priori construct or model of curriculum was used to define what constituted a differential coursework pattern. Rather, patterns were viewed as aggregations of (a) individual courses, (b) combinations of courses taken concurrently and (c) sequences of courses taken over time. The smallest unit of analysis for a coursework pattern was a single course for all the participating institutions with one exception. At Evergreen State College, the smallest unit used in the analysis was the programs; the nature and characteristics of Evergreen's academic programs is described more fully in Section 5 of this report. For the purposes of analysis, each program examined had 9 attributes represented by the residual item-type scores of students enrolling in the program. Programs with sufficient enrollment by the student sample were grouped according to the collective item-type scores of the students enrolling in the program; thus, each program examined had a mean residual score gain for each item-type.

The effect of individual programs, combinations of programs and sequences of programs on test score residuals was determined using hierarchical cluster analysis. The secondary validity of the program clusters (patterns) was derived and their relationship with the 9 item-type residual scores was determined by

11

discriminant analysis. The resultant patterns were then examined in terms of their role in the formal general education structure of the institution, the dominant type of instruction represented in specific patterns, and the nature of learning represented in those programs.

The basic steps in the DCP cluster analytic model were as follows. Coursework patterns were derived from the transcripts of graduating seniors at each of six postsecondary institutions and program patterns were determined for Evergreen State College students. First, student achievement was determined by removing the predicted effect of SAT sub-scores on GRE item-type scores; the 9 item-type residuals served as 9 criterion variables of general learned ability during the student's baccalaureate years. Second, the mean residual scores of students enrolling in each course were determined and attributed to that course. Evergreen State students enroll in curricula called "programs" which may involve study over more than one term or with more than one subject. Programs were the units of analysis at Evergreen State. Third, programs were taxonomized (or grouped into patterns) using one of two cluster analysis procedures according to the 9 GRE residual scores of the students. Fourth, the secondary validity of the clusters and the contribution of each of the 9 item-type residuals to the clusters were established by discriminant analysis. Fifth, the resulting patterns were described in terms of (a) the combinations and sequences in which the students enrolled in them, (b) the predominant curriculum and instructional characteristics of the programs in the resulting pattern, and (c) disciplines and general education requirements represented in the programs.

Following these procedures produced a set of hypothesized relationships between program patterns and measures of general learned abilities. To provide primary validation of these relationships, a second sample of graduating seniors was drawn from each participating institution, and the procedures of the cluster analytic model were repeated to determine the extent to which the relationship

of the coursework patterns or program patterns with the criterion measures of general learned ability were replicated. This report presents an analysis of the combined two samples of Evergreen State College graduating seniors.

The conceptual framework for the cluster analytic model was initially derived from a review of selected literature. This review revealed that no one curricular model and no one analytical process clearly identified the effect of differential coursework on general learned abilities. Development and testing of an analytic model for studying the differential effect of coursework in a complex curricular organization was required. A conceptual-empirical approach was adopted for the examination of the research problem, wherein a conceptual framework of student course decisions and selections guided the empirical search for coursework patterns associated with gains in general learned abilities.

During the first DCP Project year (1986-87), a model was postulated to determine effects associated with differential coursework patterns on selected measures of general learned abilities of students. Requirements of this model were that it should have explanatory power regardless of institutional setting and should function independently of the particular assessment tool (i.e., the GRE exams). The generalizability of the model to other institutional settings was tested in the second year of the project. The model as initially postulated was refined using a supplemental historical data set of student transcripts and test scores from Georgia State University (GSU). The preliminary testing and analysis of the model occurred later in the first project year, when the first sample of graduating seniors at GSU (Sample Group #1) was given the GRE and LSI instruments. Results from that administration were analyzed in previous reports (Ratcliff, 1988, 1989).

During February 1988, a random sample of graduating seniors at the institutions participating in the study (Evergreen State, Ithaca, Georgia State, Mills, and Stanford) were given the GRE and LSI. Sample Group #2 at GSU was

13

given the GRE, LSI and MPDQ in February 1989. Educational Testing Service scored the GRE tests, prepared a data tape of score results, and retired and released the test forms to the five institutions tested in May 1988 for Sample #1 and in May 1989 for Sample #2. This report presents the findings of the application of the DCP cluster analytic model to the tests and transcripts of a combined sample of 1987-88 and 1988-89 graduating seniors at Evergreen State College.

Organization of the Report

This report is divided into 7 sections. Each section describes a major component of the research activity. Section 1 describes the theoretical basis for the differential coursework hypothesis, the development of the cluster analytic model as a means of testing it and the research objectives of the DCP project. Section 2 provides a detailed description of the procedures and methodology used in conducting the research. The cluster analytic model is described procedurally. Section 3 portrays the major characteristics of two samples of graduating seniors whose transcripts and test scores were examined as part of the data gathering and analysis. Section 4 describes the characteristics of the programs found on the student transcripts in terms of the criterion variables: the GRE item-type residuals. Section 5 depicts the intended curriculum of the institution, its goals, curriculum organization and structure. Programs found on the student transcripts are compared with the general education requirements of the institution. Section 6 reports the findings from the combined sample using the qualitative cluster analytic procedures described in Section 2. Section 7 offers a summary, conclusions and recommendations emanating from the research.

Acknowledgments

This report is the result of many people's efforts. First and foremost, Steve Hunter, Director of Institutional Research at Evergreen State, served as liaison to the project. He arranged for the compilation of Evergreen student transcripts and explained the project to faculty and students often wary of anything involving standardized testing. He arranged for the testing and managed the retest administration when a snowstorm inte-vened on our carefully-laid plans. Steve also coordinated the merger of test and transcript data which ultimately served as the basis for the current study. This he did with the support and counsel of Carolyn Dobbs, Academic Dean and Patrick Hill, then Provost of Evergreen State.

At Educational Testing Service, Craig Mills and Susan Vitella arranged for and oversaw the special administration of the GRE and the preparation of data tapes that reported the results of the study. At Iowa State University, where DCP w.s first housed, Marva Ruther was project secretary, handling the many details in the development of the initial reports and analysis of data. Robert Strahan and Larry Ebbers were the DCP Project team members who oversaw the actual administration of GRE testing at Evergreen. While many were involved in the production of data upon which this report is based, only the authors are responsible for the findings and conclusions reported here.

15

I. What Are General Learned Abilities?

The cluster analytic model does not present nor does it rely on a particular theory of student learning. There is widespread disagreement on what constitutes "general learned abilities", and that disagreement is manifest in the variety of general education goals and degree requirements found in American higher education (Bergquist, Gould & Greenberg, 1981; Carnegie Foundation for the Advancement of Teaching, 1979; Gaff, 1983; Levine, 1978). Within the term "general learned abilities", we mean to include such frequently used terms as "higher order intellectual processes" (Pascarella, 1985), "academic competencies" (Warren, 1978), "generic competencies" (Ewens, 1979), "generic cognitive capabilities" (Woditsch, 1977), and "general academic ability" (Conrad, Trisman & Miller, 1977). Disagreement on terminology is but one aspect of the problems associated with measuring the general learning of students as undergraduates.

Current notions of how to assess college outcomes call for multiple measures of student achievement. No one measure has been found to accurately reflect the variety of definitions of general learning and cognitive development, the mixture of curricular goals and institutional characteristics found across the landscape of higher education and among the diversity of instructional procedures and curricular organizations of undergraduate higher education. The result has been a call for multiple measures of assessment of student learning. Policymakers and academic leaders tend to believe that since colleges and universities have broad missions and goals, assessments should be comparably broad enough to provide evaluation about as many institutional intents as possible (Loacker, Cromwell & O'Brien, 1986; Nettles, 1987).

Given the variety of terms, intents and theoretical frameworks used to explain "general learned abilities", the research design presented here is

criterion-referenced. That is, the design is not based on any one notion of what constitutes "general learned abilities" or any one college curricular structure intended to promote student cognitive development. The design permits the use of multiple and different measures of student outcomes, although the design is being developed and tested using one set of such measures. The validity of the assessment of student learning within this research design, then, is dependent upon the validity of the outcome measures selected, rather than the degree to which the student outcomes suffice a global (goal-free) measure of student learning. The design fulfills the need for multiple measures and criterion-referenced measures of student learning. However, because the design described below is not dependent upon any given college curricular structure or organization and, in fact, will be tested in five very different higher education institutions, it is free of bias engendered by specific institutional goals.

What constitutes student achievement?

Another question encountered in determining what constitutes general learned abilities is that of what composes the gains resulting from a college education. Simply measuring how graduating seniors perform on a series of tests is not a sufficient basis for generalizations about the effect of college on student achievement. First, the assessment of student outcomes is heavily effected by the students' academic achievement prior to entering college (Astin, 1970a, 1970b; Bowen, 1977; Nickens, 1970). In fact, standardized tests used for college admission, such as the Scholastic Aptitude Test (SAT), have been shown to be strongly correlated with tests used for graduate and professional school admissions, such as the General Tests of the Graduate Record Examination (GRE). These correlations have been demonstrated for the total and sub-scores on the two tests, suggesting that a large proportion of what postcollege tests, such as

17

the GRE, measures are attributable to student learning prior to college. Since the SAT and GRE were used in the development of the DCP Cluster Analytic Model, further examination of their measure of general learned abilities follows.

Nichols (1964) studied the effects of different colleges on student ability as measured by the Graduate Record Examination's Aptitude Test. A sample of 356 National Merit finalists attending 91 colleges was used. The effects of college characteristics, major fields and types of colleges on student GRE scores were examined while controlling for the precollege student characteristics. Strong correlations existed between the students' SAT scores and the students' GRE scores which ranged from .65 for the Verbal to .76 for the Quantitative. The results from this research indicated that the college students attended did have an effect on their GRE performance. There was a significant tendency for colleges to separate Verbal & Quantitative scores and raise one while lowering the other. Therefore, "the effect of college appears to be one of directing the students' abilities into verbal or quantitative channels rather than affecting the overall level of ability" (p. 52).

Rock, Centra and Linn (1970) and Rock, Baird and Linn (1972) examined SAT and GRE area test scores of 6,855 students who graduated from ninety-five colleges, predominantly small, private liberal arts institutions. The correlation between college means on SAT-Verbal and GRE-Total was 0.91. The fact that the standardized precollege and postcollege tests, such as the SAT and GRE, are strongly correlated should not be surprising. Students typically bring 12 or more years of formal education with them upon entrance to college, and since the college years traditionally constitute 4 or 5 years, a large proportion of general learned abilities of students should be attributable to learning experiences prior to college admission.

The strong relationship between the SAT and the GRE is both an asset and a liability. The use of the SAT subscores as precollege measures and the GRE

18

item-types as postcollege measures __does__ provide a basis for controlling the effects of student academic achievement using comparable definitions of general learned abilities and comparable testing procedures. However, the strong correlation between the two tests leaves only a small amount of explained variance between precollege and postcollege scores to attribute to general learning associated with a baccalaureate program.

The dilemma posed by the use of the SAT and GRE as measures of general learned abilities is exacerbated by the student population differences upon which the tests are normed. Adelman (1985) estimated that 25%-30% of graduating seniors take the GRE General examinations, while 60% of the graduating high school students take the ACTs or SATs. These rough percentages are understandable since fewer individuals choose to continue their education from bachelor's to graduate study than do those who choose to go to college from high school. Nevertheless, a consequence is that the GREs are normed on a higher ability population than the SATs (Pascarella, 1985). The individuals taking the GREs constitute a self-selected sample, driven in part by the requirements of graduate schools, professional schools, departments offering graduate degrees, and organizations requiring such examinations as part of the formal application for fellowships and scholarships (Adelman, 1985). In the DCP project, a concerted effort was made to have each sample of graduating seniors not limited to this self-selected group who would have taken the GRE as a normal event toward their application to graduate school. DCP project sampling procedures are discussed in Section 3. Nevertheless, it is important to note that the Graduate Record Examination can be accurately viewed as a measure of a student's predicted performance in graduate school as well as a measure of that student's general academic accomplishments as an undergraduate.

The GRE and SAT tests have been criticized for (a) the bias resulting from groups upon which they were normed (Adelman, 1985; Nettles, Thoeny & Gosman,

1986) and (b) their limitation in measuring higher order reasoning skills.
These criticisms notwithstanding, the GRE and SAT tests do provide an
economical, practical, and valid way of measuring selected general learned
abilities while controlling for the incoming academic accomplishments of
freshmen (Astin, 1968; Hendel, 1977). Critics of the GRE and SAT as measures of
general learned abilities attack the validity of the measures themselves. These
criticisms primarily are based on the use of sub-scores and total scores of the
tests. The use of either the GRE or SAT as multiple measures of general
learning previously had not been widely explored (Adelman, 1988). The DCP
project provided a detailed assessment of GRE item-types as discrete measures of
general learned abilities.

While the GRE and SAT tests were used in the development of the DCP cluster
analytic model, it is not dependent upon these sets of measures. The model
could be employed using another set of correlated precollege and postcollege
tests in a longitudinal analysis (for example, see Pike, 1988). For pragmatic
and economic reasons, the cluster analytic model was developed using SAT and GRE
tests. The model was initially developed using the two sub-tests of the SAT and
GRE: the SAT Verbal Test (SAT-V), the SAT Mathematical Test (SAT-M), the GRE
Verbal (GRE-V), and the GRE Quantitative (GRE-Q). Subsequent development and
testing of the model employed the 9 item-type parts of the GRE as multiple mea-
sures of student learning. Therefore, the cluster analytic model will be des-
cribed in terms of these 9 measures of student general learned ability.

Prior research suggested that the 9 item-type scores were independent
measures of general learning. Wilson (1985) examined the criterion-validity of
the 9 item-type part scores of the GRE General Test to the prediction of
self-reported undergraduate grade point average (GPA). For his research, Wilson
used the GRE scores of 9,375 examinees in 9 different fields of study
representing 437 undergraduate departments from 149 colleges and universities.

Data were first standardized within each department, then pooled for analysis hy field of study. Results suggested that the GRE item-type scores did differentiate undergraduate GPA by field of study. This research and other studies (Powers, Swinton & Carlson, 1977; Swinton & Powers, 1982; Wilson, 1974) indicated that the 9 item-type subparts of the GRE measure different and somewhat unique general learned abilities. Initial testing of the DCP model at four colleges and universities also indicated that the item-types constituted independent measures of generai learned abilities (Ratcliff, 1988, 1989).

The GRE General Test consists of three sections: verbal, quantitative and analytical; within each test section are specific types of test items (i.e. Verbal: analogy item; Quantitative: quantitative comparisons; Analytic: logical reasoning). There are 9 item-types within the General Test; the residual differences between the observed GRE scores (postcollege measure) and the GRE scores predicted by the students' corresponding SAT scores (precollege measure) were used to gauge general learned abilities attributable to the students' undergraduate education. The residual differences between the observed and predicted values of each GRE item-type served as the 9 measures of student gains in general learned abilities during the time in which the student was enrolled in college. Thus in an economical and practical way, these item-type residuals represented a set of multiple measures of general learned abilities which accounted for and controlled the effect of precollege student achievement on postcollege student outcomes.

What constitutes a coursework pattern?

The prevalent way to view the college curriculum is by its intentions, rather than by its results (Warren, 1986). Since measuring the effects of the curriculum is problematic, it is not surprising that many studies presume rather

21

- 6 -

than test the effect of different patterns of coursework.

The college curriculum is substantiative, additive and temporal. In terms of cognitive theories of curriculum development, both content and process contribute to developmental learning in students (Tyler, 1950; Taba, 1962). Essentialist and constructionist theories of curriculum stress combinations of subjects (core curricula, great books, etc.) as influential on general learned abilities of college students (Fuhrmann and Grasha, 1983). The medieval university curriculum was organized according to combinations and sequences of courses as well as individual subjects (Rudolph, 1977); the seven liberal arts were sequenced into the prerequisite subjects of the quadrivium (arithmetic, geometry, astronomy, and music) and the higher order subjects, the trivium (logic, grammar, and rhetoric). Together, the quadrivium and trivium provided an individual with the general learned abilities needed to study the three philosophies of Aristotle: natural philosophy (physics), moral philosophy (ethics), and mental philosophy (metaphysics). These combinations and sequences of coursework have been generalized more recently into concepts of breadth and depth as criteria by which to describe higher education curricula (Blackburn et al., 1976).

While the notion that combinations of concurrent coursework and that developmental sequences of coursework lead to effects in the general learned abilities of students is derived from the medieval university, it is underscored and further supported by the research of contemporary developmental theorists. Perry (1968) for example, stated that development "consists of an orderly progression of cognition in which more complex forms are created by the differentiation and reintegration of earlier, simpler forms" (p. 44).

The value of curricular substance and sequence are presumed in formulations of core curricula, in the four levels of study (freshman, sophomore, junior and senior years), in the corresponding practice of assigning course numbers

22

according to those divisions, and in the practice of assigning course

prerequisites. To assess the impact of these coursework patterns on the general

learned abilities of students, the additive, substantiative and sequential

characteristics of student course-taking need to be examined. These notions of

what ought to be taught and what students ought to learn presumably represent

the philosophical and educational aims of the particular college.

Nevertheless, a distinction should be made between those patterns of

coursework intended to fulfill undergraduate program and degree requirements and

those patterns of coursework which students actually choose (Boyer & Ahlgren,

1981, 1982, 1987; Warren, 1986). Intentional patterns of coursework are

provided in a variety of publications issued by the institution: the college

catalog, the annual schedule of times and days of courses, and program descrip-

tions issued by departments and divisions within the college. Richardson et al.

(1982) provide evidence that a minority of students may consult these statements

of curricular intent prior to making decisions about which courses to choose.

Other forms of intentional coursework patterns are the lists of courses or sub-

jects required for certification or licensure in a particular profession, occu-

pation or technical field. Such lists of coursework may be compiled by practi-

tioners and academics of a given discipline or profession to accredit college or

university programs. Just as the curriculum of a particular college may repre-

sent the philosophy and educational aims of that institution, so too may the

certification, licensure and accrediting standards articulate the intentions of

state, regional, disciplinary and programmatic associations. All are intended

patterns of coursework in the curriculum whose measure of effectiveness, in

part, is the extent to which these patterns accomplish their aims in practice.

In a college curriculum, a single course may be the smallest unit of

analysis. In assessing the impact of the curriculum on the general cognitive

development of students, the course constitutes a datum in the analysis. A

23

- 2 -

pattern of data is a design resulting from "the relation among a set of objects" (Romesburg, 1984, p. 278). In this case, the objects are courses. Therefore, a coursework pattern is a design resulting from relationships among courses. A cluster of courses (objects) is a set of one or more objects found to be similar to one another according to a given set of attributes. In the DCP project, courses were grouped according to the extent of gairs in general learning of the students enrolled in the courses. Thus, for the purposes of the DCP cluster analytic model, a <u>cluster of courses</u> is used to denote a pattern of coursework with an empirically derived set of relationships. Stated another way, a cluster of courses is a pattern based on the actual enrollment of students, rather than the intended enrollment pattern of the college or university. This distinction is important in order to differentiate between the consequences of the college curriculum and its intents.

Sources of data for coursework patterns

Arguments about what is and what is not an effective college curriculum are for the most part based on seasoned speculation, nostalgia about academic traditions, and unrealistic expectations of curricular coherence among and within the over 3,000 colleges and universities in the United States (Conrad, 1986). In most instances, the data used in describing the status of general education are derived from catalog studies and enrollment analyses. These data may not present an accurate picture of general education as it functions in students' programs.

Catalog data provide evidence of trends in offerings but ignore student behavior. Catalogs indicate which courses a student should take to fulfill degree requirements and describe the intended contents or outcomes of the courses; collectively, this compendium of courses and articulation of degree and program requirements ideally leads to the accomplishment of the educational

24

philosophy and curricular goals of the institution. White (1979) claimed that "college catalogs generally have little to do with reality", and that "the educational ideas expressed ... are subsequently neither perceived nor accomplished" (p. 39). The ideals expressed are important, "yet when subjected to the rigorous scrutiny of time and experience, the academic promise is often not realized" (p. 39). White suggested that all collegiate programs, including general education, should be assessed in light of their original claims and promises.

Dressel and DeLisle (1969) investigated college curriculum trends using catalogs as primary sources of information. They conceded that data derived from studies like theirs must be interpreted circumspectly, asserting that:

> ... ambiguities and contradictions arise in the use of catalogs for research of curriculum practices, because what appears in the catalog as policy is in reality often left to interpretation of individual advisors and individual departments, and what is in reality required by an individual department is often not stated as policy in the catalog (p. 75).

> ... the limits described by departments tend to be more exacting and demanding than those stated in the institutional requirements listed by the college. Thus, a question arises as to whether each student actually has flexibility and innovation claimed in the general statements or whether the department control of the major serves as a limiting and inhibiting factor in this respect (p. 78).

Dressel and DeLisle also noted that catalogs, as sources of information about the curriculum, cannot be examined to determine the extent to which curriculum policies and statements correspond to actual course-taking practices of students. Because of ambiguities, inaccuracies, discrepancies, and omissions in wording in catalogs, the interpretation of the catalog varies significantly among both students and advisors. Also, catalogs often do not present a rationale for course requirements, nor is there a way to determine how, why, or when requirements were introduced. Likewise, there is no way of determining "whether claimed articulation of liberal with professional education and of breadth with

25

depth has been successfully achieved" (p. 79). In sum, catalogs do not provide information about the consequences of the coursework on student learning, and therefore, are not appropriate for learning outcomes research.

Assessments of the outcomes of college have little meaning unless compared with either a normative group of students or the intended general educational goals of the institutions. Catalogs are often the most comprehensive statements of the intended substance and sequence of intended learning activities. From our standpoint, they provide a basis for comparison of college intentions with college outcomes. The DCP project closely examined the extent to which the college catalog represented the formal curriculum students experience in college. This comparison is presented in Section 5 of this report.

While college catalogs are inadequate in describing course-taking behavior of students, enrollment figures as data sources also provide limited inform-ation. Although such figures are often used for evidence of curricular trends among undergraduates, enrollment analyses do not describe the actual course pat-terns of enrollees; however, they do reveal the extent to which different seg-ments of the college population choose particular majors or general education courses (intended coursework patterns). Noting the inability of enrollment analysis to determine student reasons for course selection, Friedlander (1979) used transcript analysis to examine the effect of changes in the composition of the community college student body on humanities enrollments. Adelman, in an analysis of the Postsecondary Education Transcript Sample (PETS) of the National Longitudinal Study of the High School Class of 1972, summarized the advantages of transcripts as unobtrusive, empirical artifacts of student learning, "transcripts neither exaggerate nor forget" (1989, p. 1).

Student transcripts as a data source

Student transcripts are a rich, unobtrusive and problematic source of information about student course-taking behavior. Warren (1975) used transcripts to determine coursework patterns among college students in a study of 50 history graduates of different four-year colleges. The student course-selection patterns in history, as revealed in these transcripts, indicated that within the discipline there were at least three or four different history programs. This finding demonstrated that although students receive similar degrees, they do not necessarily have the same educational experiences. Warren's study suggested that students shape their own curriculum as they exercise options in choosing courses to complete credit hour requirements. Furthermore, Warren demonstrated that transcripts could be used to discern broad curricular patterns.

Blackburn and associates (1976) used transcript analysis in their investigation of curricular change and course-taking behavior in two- and four-year colleges and universities between 1967 and 1974. One of the goals of the study was "to determine how students utilize elective time" (p. 20), and, since some two-year transcripts did not indicate institutional requirements for both general education and the major, the researchers could not ascertain what courses were elective and which were prescribed. Consequently, the transcripts from two-year colleges were eliminated from the sample used to describe student course-taking patterns. The question of treatment of transfer students in transcript studies and college impact studies has been a consistent methodological problem (Astin, 1970a).

Prather and associates (1976) used transcript analysis to study undergraduate grading practices at Georgia State University and investigated differences in grading patterns by major fields of study while controlling for such antecedents as scholastic aptitude, demographic background, course types, and

27

longitudinal trends. They found that major field was strongly associated with the grades students received in courses throughout the curriculum. This research and previous grade studies supported the proposition that the various parts of the curriculum have different grading standards, arguing against the use of GPA as a proxy measure of general learned abilities in college students. The studies by Warren (1975) and Blackburn and associates (1976) used small samples of transcripts (5 percent) due to the time demanded in reading and assessing all the courses on each transcript. Prather and associates, however, used an electronic database of student transcript information to examine an institutional cohort; use of electronic databases of records enabled the researchers to examine larger samples of student records, thereby permitting analysis of larger sections of the curriculum.

Transcript analysis has been used to examine the general education component of the undergraduate curriculum as well. The dean of instruction and curriculum planning at the University of Pennsylvania used transcript analysis in an effort to determine which courses among the many listed in the college catalog were actually selected by arts and sciences graduates (Carnegie Foundation, 1979). He found that 1976 graduates of arts and science programs had selected "a core of 29 courses" (p. 97) in the curriculum. However, not all students chose the same combination of courses, and "many of the thousands of courses in the catalog that were not included in the core list were found on individual transcripts" (p. 97). This study illustrated one of the persistent problems in using transcript analysis to identify course-taking patterns: the enormous range of possibilities of course sequences generated by student choice in a large, multi-purpose university. It also suggested that, for whatever reason, there is a limited number of courses which most students select to complete the general education requirements of the undergraduate program.

29

Beeken (1982) used transcript analysis to examine the course-taking behavior of a sample of students in three Virginia community colleges. The purpose of the study was to determine the number and types of general education courses selected by students to meet the general education requirements of the Virginia Community College System. The study did not confirm the conclusion of the Carnegie Commission that the general education curriculum was a "disaster area", although the programs of many students did not present a balance of disciplines; students apparently minimized the number of mathematics and science courses in their program of study. Both those who completed an associate of arts degree and those who did not exceeded the minimum requirements fr~ general education courses. The number of courses taken in different curricular areas of general education were related to enrollment status, age, and sex.

One of the largest collections of student transcripts is the Postsecondary Education Transcript Sample (PETS) on the National Longitudinal Study of the High School Class of 1972 (NLS). PETS data consists of 22,600 student transcripts. While NLS has several precollege measures of achievement (high school grades, SAT, etc.) and the coursework selected by students who attended college is represented, the NLS data has no available post-baccalaureate measure of general learning. Adelman (1989) used NLS/PETS and the NLS 5th Follow-Up Survey to demonstrate relationships between coursework taken in community colleges and success in attaining bachelor's and advanced degrees, career aspirations and plans, and self-reported attributes of the jobs the students held 15 years after high school graduation. In this analysis, transcripts proved to be a powerful, non-obtrusive measure of the relationship between what a student planned, what they studied at college, and what the nature of their work was a decade and a half later.

In a limited number of studies, transcripts have been found to be a useful, valid and reliable source of information on student course-taking behavior.

29

They provide evidence of the combination, sequence and performance of students in the patterns of courses in which they enroll. As archival records, transcripts are unobtrusive data. While most studies have limited their use of transcripts to a manual examination of a small sample of student records, there is evidence that such records, stored on a college or university computer, can be used to examine the course-taking behavior of a whole class, cohort or population of students. The cluster analytic model for determining the associated effects of coursework patterns on the general learned abilities of college students uses transcripts in precisely this manner. Transcripts maintained on an electronic database can be merged with student score residuals for the purposes of assessing the effects of the curriculum on student learning.

A conceptual framework for analyzing coursework patterns

The differential coursework hypothesis posits that students who enroll in different coursework will show different levels or types of gain in general learned abilities (Benbow and Stanley, 1980, 1982, 1983; Pallas and Alexander, 1983). While all the courses Student X chose collectively affect X's gains in general learning, the effects of an individual course on X's transcript may vary in its contribution to such an effect. The effect of individual courses may be mediated by prior student aptitude, ability, achievement, and interests.

For the purpose of analysis, the effects associated with a particular course is proxied by the residual score of the students who enrolled in that course. A pattern of data is a design resulting from "the relation among a set of objects" (Romesburg, 1984, p. 278). In this case, the objects are courses and at Evergreen State College the objects are programs. Therefore, a coursework or program pattern is defined here as a set of courses or programs having comparable effect on one or more student residual scores. The differential coursework hypotheses is rejected when no patterns are discernible

30

among the data--when the residual scores of students are uniformly attributable to enrollment in any and all courses in the curriculum. The hypothesis is affirmed when students who perform well on one or more postcollege measures tend to enroll in certain courses or programs and not others.

At this point in the inquiry, the DCP cluster analytic model is not concerned with reasons for the effect. Rather, the next step is to identify and classify courses according to the score gains of students who enrolled in them, regardless of the factors that may have brought the students to enroll in the courses. Also, the model is not yet concerned with characteristics of the students, although those characteristics may covary with the course selection or achievement variables (Elton and Rose, 1967; Prather et al., 1976). Thus, in the examination of the effect of course patterns, there is no implication of direct causality of course patterns upon achievement (Astin, 1970a, 1970b).

There is reason to presume that the effect of a single course may vary according to what place it holds in the pattern of courses a student chooses (Prather et al., 1976). For example, if courses at a particular college are sequenced according to level (e.g., 100 level courses are intended for freshmen and 400 level courses are intended primarily for seniors), the effect of History 101, "Survey of Western Civilization", may differ for Student X who enrolls as a first term freshmen from Student Y who enrolls as a final term senior. Conversely, logic holds that the effect of History 451, "20th Century American Foreign Policy", may differ for the first term freshman and the last term senior (Rudolph, 1977; Veysey, 1973). If a course is viewed as contributing to the residual score for a particular measure of general learning, then a course's effect may vary according to its place in the student's pattern of courses. Therefore, course sequencing should be considered in the examination of course patterns (Bergquist, 1981).

31

Likewise, the effect of a particular course may be associated with the effect of other courses in which the student may be concurrently enrolled. Richardson et al. (1982) and Roueche and Snow (1977) noted that students may be advised to enroll in elementary writing or mathematics courses concurrently with other courses requiring the basic skills these elementary courses teach. Under such practices, the student may have much less chance to succeed in college. Traditionally, the combination of courses in which a student enrolls within a given term is presumed to have effect (Bergquist, et al. 1981; Rudolph, 1977; Veysey, 1973) and therefore also should be considered in the analysis of course patterns.

Thus, there is a great deal of uncertainty associated with why a particular student chooses a particular course at a particular time in his/her program of study. A poor grade in "Trigonometry" may cause Student Y to select a remedial mathematics course over "Introduction to Calculus". Student X, who received a high grade in "Trigonometry", may not enroll the following term in "Introduction to Calculus" because the time it is offered conflicts with a course Student X is required to take within his/her major. Or the Calculus course may be filled when Student X tries to enroll. Many factors shape the combination of courses a student chooses in a given term and the sequence of courses are represented across terms in the transcript.

A modern research university may present 2,500 to 5,000 undergraduate courses from which a student may choose 35 to 45 courses to complete the baccalaureate degree. Each semester or quarter a student enrolls, that student selects several courses. Each term of registration represents a stage in the overall decision-making process which generates the patterns of coursework found on the student transcripts at time of graduation. Each enrollment decision is limited and shaped by those courses in which the student has previously enrolled and the various degree requirements and prerequisites that are enforced during

32

the registration process. At each successive decision-point, the student is progressively more immersed in the college environment, the norms and values of the student's peers, and the norms, values and expectations of the subjects the student selects to study.

The analysis of the pattern of courses a student chooses is a sequential decision-making process wherein certain conditions exist:

1. students make course selections in an environment of uncertainty about the consequences of the choices;

2. there are multiple reasons why students enroll in each course;

3. there are multiple options available to the student at each decision-point (term registration period);

4. student course selections are sequential; there are different decision-points (terms) in which parts of the coursework pattern are chosen, with prior decisions having some bearing on future decisions.

Under the conditions listed above, students may choose courses to minimize uncertainty and risk (i.e., seek what they perceive to be "easy" courses). They may also seek courses which will maximize the efficiency (i.e., fulfill degree and graduation requirements with a minimum amount of time), or maximize effectiveness (e.g., "it's a hard course, but I need to pass it if I'm going to major in engineering"). In this way, the succession of registration decisions comprising the student's pattern of coursework conceptually represents a multiple-stage, decision-making process (Buchanan, 1982; Bunn, 1984).

According to Pace (1979), one variable in student development is the amount of time and effort invested by the student. This premise, that student involvement in learning advances student achievement, guided the recommendations of the NIE Study Group's Report on Conditions of Excellence in American Higher Education (1984). Not only the kind and quality of cognitive activities in which the student engages, but also the level of effort exerted by the student in understanding and using the knowledge and abilities gained influence the

33

quality of student learning. The student's effort in courses is "impressed" (Pace, 1979) by attitudes of the perceived usefulness of the course and the perceived difficulty of the course. These perceptions influence the kind and quality of student investment in learning. Coyne and Lazarus (1980) found that such investment involved both cognitive and subjective elements, leading to whether the experience is viewed as a challenge or a threat. The perceived difficulty of courses influences student enrollment decisions and thereby contributes to the multiple-stage enrollment decision-making process through which the student compiles his or her particular collection of coursework.

In summary, the literature suggests a number of possible interactions between student and curriculum each time a student makes course selections. The effect of courses on general learned abilities may vary according to the course itself, the time of enrollment in the student's baccalaureate program, the concurrent or sequential relationship to other courses in which the student enrolls, the predominant learning style of the course and of the student, the curricular design of the course, and the risk-taking behavior the student exhibits at each enrollment decision-point. The DCP cluster analytic model calls first for the identification of student achievement (i.e., student score residuals), second for the classification of courses found on student transcripts into patterns according to their associated effects on the student residual scores, and thirdly, for the further classification of courses within each identified pattern according to their sequence and combination and the common curricular characteristics of courses found within a given pattern of coursework. The model provides a basis for examining the extent to which the empirically-derived patterns of coursework reflect institutional mission and curricular goals, general educational requirements, the values, norms and mode of inquiry represented by the disciplines studied, and the demographic characteristics of the students. The model accomplishes these objectives

through the use of cluster analysis, a statistical procedure which has been used throughout the physical and social sciences to derive empirical taxonomies of objects in a variety of settings. Cluster analysis, since it has been infrequently employed in education, is described in greater detail in following Section 2.

Research Objectives of the DCP Project

The objectives of the DCP cluster analytic model are:

1. To determine student academic achievement in general learned abilities gained during the baccalaureate program. This achievement was measured by the fixed criteria of the residual scores on the 9 item-type subparts of the General Test of the Graduate Record Examination (GRE), once the effect of precollege achievement as measured by Scholastic Aptitude Test (SAT) scores were removed.

2. To classify the coursework taken during a student's baccalaureate program according to its associated effects on the student's general learned abilities, as measured by the GRE. This coursework was determined by a cluster analysis of student transcripts wherein courses will be described and classified into patterns according to the GRE residuals of the students enrolling in them.

3. To test the secondary validity of the coursework clusters and to identify outlying cases within each cluster of courses, discriminant analysis was applied to the results of the cluster analysis. Through examination of pooled within-group correlations of discriminant functions with GRE item-types and the cluster group means on each discriminant function, relationships between cluster groups and item-type residual scores will be determined.

4. To describe the resulting patterns of coursework according to: sequences and combinations of courses within the cluster, according to term of enrollment data found on students transcripts, and the common curricular characteristics of the institution.

Samples used in the initial development
of the cluster analytic model

For the purposes of building and testing the cluster analytic model, an historical database was developed at Georgia State University. This database consisted of 1,024 students who began their baccalaureate education at GSU,

graduated, and then continued in graduate or professional education at GSU. To qualify for inclusion in this database, a native student must have completed 14 or more quarter credits. This same criteria applied for Clayton State students. A student who completed more than one quarter (15 credits) at another institution prior to enrolling at GSU was not included in the Historical Group. The database was drawn from all student transcripts at GSU between 1975 and 1985. This population was selected because: (1) it was a readily available database, (2) prior research demonstrated that it was amenable to statistical analysis (Prather & Smith, 1976a, 1976b; Prather, Smith & Wadly, 1976), and (3) the transcript records also contained SAT and GRE information as well as courses taken.

From this database, 56 student records were found to contain the Scholastic Aptitude Test verbal scores (SAT-V), Scholastic Aptitude Test mathematical scores (SAT-M), Graduate Record Examination verbal score (GRE-V) and Graduate Record Examination quantitative score (GRE-Q). All student identification information was removed from this database at GSU, so the individual identity of the student was unknown to the researchers developing the cluster analytic model. It should be noted that these 56 student transcripts were representative of the GSU student population in every way other than by major. Approximately 20% of the sample were found to be psychology majors and another 20% were English majors. The spread of SAT scores appeared to otherwise approximate the demographic characteristics of GSU students. Therefore, the sample did provide a database in order to develop the model.

The historical database contained 1,024 transcripts upon which were listed an unduplicated count of 2,470 discrete course numbers and grades. The sample of 56 transcripts with complete GRE and SAT test score information contained an unduplicated count of 1,065 discrete course numbers and grades. No evidence

existed that a particular course offered in a particular year was indeed identical to a course bearing the same course number in another year. The comparability of courses in a cross-sectional study of a single cohort of college seniors may be less than that of a historical group, since the potential differences between courses bearing the same course identification number would only vary over about 4 years (from courses taken by the cohort when they were freshman to those completed as seniors). A historical database accentuates the potential for significant changes in course structure, content or staffing. Nevertheless, for the purposes of model building and preliminary analysis, courses of the same course number taken in different semesters or years were assumed to be comparable.

Courses repeated for credit were eliminated from the analysis. For example, MUS 101 was found to be a performance music class. One section of this class might be performance oboe, while another might be performance piano. Thus, students interested in music enrolled in multiple sections of the class during one term and enrolled repeatedly in the course over several terms. Likewise, HON 326 was found to be an honors seminar in the arts and humanities one quarter, in the social sciences the next quarter, and in the physical and life sciences yet another quarter. Therefore, these courses were eliminated from the analysis because they violated the assumption of comparability of the course number over the quarters represented by the historical database.

37

Figure 1-1. Transcripts in the GSU Historical Database.
==

| | Total Database | | Sample | |
	N	Percent	N	Percent
Transcripts	1,024	100%	56	5.47%
Unduplicated Course Numbers	2,470	100%	1,065	43.12%
Courses taken by 5 or more students			101	4.09%

==

Prior research (i.e., Blackburn et al, 1976; Drees, 1982) used a five percent sample of transcripts upon which to base generalizations regarding the undergraduate curriculum. The above analysis of the Historical Group database at GSU suggested that a 5 percent sample of transcripts would yield over 40% of the available courses in the college curriculum, but only about 5 percent of the curriculum will be represented by 5 or more students in the sample. Therefore, generalizations about specific courses across a broad spectrum of the curriculum cannot be made based on the course-taking behavior of 5 or more students. Only those courses most frequently chosen by students may be included in such an analysis. Obviously, average class size has a bearing on the number of courses available for analysis, given a specified transcript sample size and minimum number of students required in each course cell in the cluster data matrix. An area needing further research is that of the relationship between sample size of transcripts and representativeness of the curriculum as a whole.

The relationship between sample size
and the college curriculum

A persistent problem in linking the undergraduate curriculum to measures of student learned abilities is the number of courses from which students may choose. As previously mentioned, students will typically enroll in 35 to 55 separate courses to complete their bachelor's degree, although the number of courses vary considerably. Students select these 35 to 55 courses from a catalog of several thousand courses at a university or several hundred at a smaller college. Linking the effect of one course to the general learning of students therefore becomes problematic.

Two samples of graduating seniors at Evergreen State College were drawn during 1987-88 and 1988-89. However, since the individual sample sizes were small, a decision was made to combine the samples together resulting in a larger sample size to examine important relationships. All of the data presented in this report concerns the combined sample.

There were 269 total programs appearing on the transcripts of the 46 students. Of the 190 unduplicated programs, 43 were taken by 2 or more students in the Evergreen State Sample. The preliminary analysis, therefore, focuses on these 43 programs (most frequently chosen by students) which represent 22.63 percent of the curriculum (as represented on the transcripts).

39

Figure 1-2a. Summary of Evergreen State Sample

	Usable Sample	
	N	Percent
Transcripts, GRE tests	46	
Duplicated Evergreen programs	269	
Unduplicated Evergreen programs	190	
Unduplicated programs taken by 5 or more students	3	1.58%
Unduplicated programs taken by 2 or more students	43	22.63%

The question of the representativeness of a sample of courses or programs
to the total curriculum is exacerbated by the lack of a precise definition of
the total curriculum. As was evidenced by the analysis of Evergreen State
student transcripts, the exact number of programs available to students does not
correspond to the college catalog. In fact, many programs are negotiated
between students and faculty and are not recurring phenomena listed in the
college catalog. The exact number of unduplicated programs listed in the
Evergreen State curriculum database was not available at the time of this
report. Likewise, the exact number of programs available for enrollment in any
given year was not available. The programs in one year were not identical to
those offered the following year. What constitutes the curriculum, in terms of
number of programs, content and variety, varies from term to term and year to
year.

Without an exact definition of the total curriculum available to
undergraduates, the representativeness of a sample of programs can only be
approximated. Since that definition of the curriculum evolves over the period
encompassed by a baccalaureate program, since students enter and exit the

4()

baccalaureate programs in different terms, and since the tenure of their undergraduate studies varies, the exact extent of courses from which a student can make choices becomes individual, nebulous and imprecise. Even so, the data from the Evergreen State Samples #1 and #2 reflect the extent of curricular change in even the most frequently enrolled programs over a six-year period generally covered by the transcripts. More research is needed in the variability of the program offerings on a yearly basis.

In the DCP project research, samples of student transcripts were drawn and those courses enrolling 5 or more students were examined in the quantitative cluster analysis for all participating institutions except Evergreen State College (where a qualitative analysis is utilized due to the low number of programs on student transcripts). When those courses enrolling 5 or more students were selected, the proportion of the total curriculum represented by those students was significantly decreased. When the initial sample size is not very large, the representativeness of the courses to the total curriculum may be seriously questioned. However, many debates regarding the vitality of the undergraduate curriculum in producing general learning among students consider only the general education portion of the curriculum, not every course listed in the catalog. From that standpoint, the representativeness of the programs included in the cluster analysis may be defined in terms of either (a) the total of programs offered during the period of enrollment of the student, or (b) the combinations and sequences of programs prescribed by the college or university to meet the general education requirements for a bachelor's degree.

The total courses or programs offered during the period of enrollment of a student is not easily ascertained at many colleges and universities. First, the student transcripts list only the courses the student chose, not what was offered that they didn't select. Student choice of coursework or programs is not made in isolation, but is made in relation to those courses/programs not

41

selected, those previously selected, and those planned for future terms.
Second, not all courses/programs offered during a given period are listed in the college catalog or bulletin. Comparing the student transcripts with the college catalog reveals that experimental courses/programs and new courses, some of which may be available only in one term or year, do not appear in the catalog. Thus, courses/programs not listed in the catalog and not selected by a given cohort of students were among the range of enrollment choices available to the students. Lastly, there are courses/programs in the college catalog which may not be given during the enrollment period of a cohort of students. While such courses/programs were not choices to the student cohort, they were regarded as part of the formal curriculum of the institution. Thus, defining the curriculum as all courses/programs available and/or advertised to a particular cohort of students may not produce an exact representation of the college curriculum. It may, in fact, obscure some of the most experimental and innovative courses/programs which, for one reason or another, did not get recorded in the college catalog.

On the other hand, if one defines the curriculum pertinent to general learning solely in terms of what general education courses/programs are required for degree completion, the distinction between what the college intends and what the effects of the college curriculum are is blurred. The possibility looms large that a student enrolled in coursework/programs that enhanced his or her general learned abilities but was not part of the formal general education requirements of the institution. Previously mentioned problems also exist with this definition of the curriculum as well: courses/programs not selected are not fully represented, courses/programs not listed in the catalog ay be overlooked, and courses/programs listed but not offered are treated as part of the range of options. In sum, the undergraduate curriculum is not a tidy item for analysis.

However, an alternate interpretation of the relationship between course/program attributes and their reliability would yield a different perspective on the above problem. In the initial analysis of the Evergreen State Sample Group, the criterion variables used were correlation coefficients between SAT and GRE scores for those enrolling in a given program. Here, the probability of error varies according to the number of enrolled students. When GRE item-type residual scores are used as program attributes, however, this does not seem to be the case. The residual scores calculated among all students exhibited high confidence levels; the regression functions of SAT scores on GRE item-type scores proved significant. The remaining concern, then, is the level of confidence attributable to a single program when it is described by the mean of student residual scores from the sample group who enrolled in the course.

At this juncture, it is important to note that the focus of the analysis is on courses or programs for Evergreen, not students. It is not the purpose of the cluster analysis to predict the population mean parameter of all the students enrolled in a program. Since the main purpose of the cluster analysis of college curriculum is to examine the effect of an unknown program enrollment pattern on student general learned abilities, the confidence level of mean residuals for an individual program is not of much importance because the attributes are in large part significantly determined by all students in the sample group, rather than by the students enrolled in that program alone. The course attributes are determined by student program enrollment pattern, not by the characteristic of a single program.

The analysis of the Evergreen State Sample #1 discussed later in this report revealed that each program cluster includes a certain variety of subjects and levels, ranging from beginning to intermediate to advanced studies. These clusters, derived of programs sorted according to the gains in general learned abilities of the students who took them, generate questions regarding the basic

43

attributes of the college curriculum: discipline, sequence and level.

Assumptions

1. One learns what one studies.

2. Courses are the primary units of learning in college.

3. Transcripts are an accurate listing of the enrollment pattern of students.

4. Most undergraduate courses are basically stable in content and instruction over time and among instructors.

5. The effects measured can be generalized to all the formal coursework in which a student enrolled.

Limitations in the analysis of curricular patterns

The analysis of program patterns that lead to higher student gains in general learned abilities should differentiate the effect of different parts of the undergraduate curriculum. The analysis should also point to that curriculum which promises to be most effective for promoting cognitive development. However, the analysis is delimited to students who were attaining the bachelor's degree at Evergreen State College and other institutions participating in the DCP Project. No analysis is presented of the differential effect of programs on those students who ended their studies prior to completion of their senior year. The following discussion refers to courses; however, at Evergreen State College programs are offered but the limitations are still applicable.

Two forms of error need to be avoided in such an investigation. One is the reductionist error of attempting to account for the variance in complex groupings of coursework through individual psychological variables. In a recent review of the literature on the effects of race, gender and class on educational attainment, Grant and Sleeter (1986) concluded that attention to one status group oversimplified the analysis of student behavior, confirming the problems

of reductionist error. The reductionist error may also occur in research equating general learned abilities within complex academic organizations with intra-group cohesion and/or with individuals' identification with an academic discipline. The study of student learning in colleges and universities is a study of student behavior _in_ such organizations, rather than the study _of_ such organizations.

A second type of error is the uniqueness-of-data approach. While it is important to acknowledge what is unique in each institutional learning environment, this should not halt the exploration of appropriate relationships of curricula within different colleges and universities. This error may emanate from the failure to conceive of these institutions as systems (1) nested within and linked to larger systems (disciplinary and professional fields) and (2) containing smaller subsystems (departments, divisions and programs) that are, in turn, linked to them (Katz, Kahn and Stacey, 1982).

Prior research suggests that student coursework patterns found to affect general learned abilities can be characterized by (1) the extraneous (other than achievement) characteristics of the students enrolled, (2) the unique or idiographic characteristics of the learning environment, and (3) the normative effect of the fields of study on learning in colleges and universities (Astin, 1970a; Pascarella, 1985). Prior research has also demonstrated that more than one model of college curriculum can explain the effect on student learning from a common set of transcript data (e.g., Hesseldenz and Smith, 1977; Kolb, 1983). Therefore the cluster analytic model identifies empirically-derived course patterns which subsequently may be examined in terms of student characteristics and idiographic and nomothetic aspects of the curriculum. In this sense, the cluster-analytic model is retro-deductive in approach and is useful to the generation of research questions and hypotheses regarding common notions of the college curriculum and its relationship to general student learning at the

45

undergraduate level.

Definition of terms and concepts

Transfer student: A student who has earned the equivalent of one or more terms
of full-time work (15 quarter credits or 10 semester credits) at another
institution of higher education prior to enrolling at the institution under
study.

Native student: A native student has obtained his/her undergraduate educational
experience primarily from the institution under study. Native students
entered the college or university with no more than 14 quarter credits or 9
semester credits earned at other institutions. Native students may have
accumulated coursework at other institutions during their bachelor's
program (i.e., a student who attends summer session at another institution
during her junior year), but such credit does not constitute a significant
portion of their baccalaureate program.

Graduating senior: A student who has declared his/her intention to graduate or
is estimated to graduate during the calendar year commencing July 1st and
ending June 30th.

Scholastic Aptitude Test (SAT) Scores: Students may have taken the SAT examina-
tions more than once prior to admission. When more than one set of SAT
scores was available for a given student, the SAT score date immediately
preceding the initiation of the baccalaureate program was used. That is,
if a student took the exam several times and entered college in September
1980, then the SAT scores from the test most immediately preceding
September 1980 were used. The SAT is a precollege effects measure, and the
more proximous the measure is to the initiation of the effects to be
analyzed, the most desirable.

Quarter calendar: The calendar usually consists of four ten-week terms.

46

Semester calendar: The calendar usually consists of two terms which average

fifteen weeks each. However, each term can be as long as twenty weeks.

Description of the GRE General Test

The GRE General Test purports to measure verbal, quantitative and analytic

abilities important to academic achievement (Educational Testing Service, 1988).

In doing so, the test reflects the opportunities and efforts of the student to

acquire these abilities.

Verbal abilities (GRE-V). One of the major subscores of the GRE General

Test is that of verbal ability. Verbal ability is described as the ability to

reason with words in solving problems.

> Reasoning effectively in a verbal medium depends upon the ability
> to discern, comprehend and analyze relationships among words or groups
> of words and within larger units of discourse such as sentences and
> written passages. Such factors as knowledge of words and practice in
> reading will ... define the limits within which one can reason using
> these tools (ETS, 1988, p. 28).

The GRE Verbal Subscore is derived from 4 types of items: analogies,

antonyms, sentence completion and reading comprehension questions. Each is

described below:

Analogies (ANA). Analogy items test students' ability "to recognize

relationships among words and the concepts they represent and to recognize when

these relationships are parallel. The process of eliminating four wrong answer

choices requires one to formulate and then analyze the relationship linking six

pairs of words" (ETS, 1988, p. 28).

Antonyms (ANT). Antonym items provide a direct test of the student's

vocabulary. However, the purpose of this item-type is not merely to measure the

student's vocabulary, but also to gauge "the student's ability to reason from a

given concept to its opposite" (ETS, 1988, p. 29).

47

Reading Comprehension (RC). To successfully complete these items, students must read narrative with "understanding, insight and discrimination". These passages challenge a student's ability to analyze using a variety of perspectives "including the ability to recognize both explicitly stated elements in the passage and assumptions underlying statements or arguments in the passage as well as the implications of those statements or arguments" (ETS, 1988, p. 31). Due to the length of the narratives around which the questions for this item-type are built, students are given ample opportunity to assess a variety of relationships, such as the function of a key work in a passage, the relationships among several ideas, or the relationship of the author to the topic or the audience.

Sentence Completion (SC). These items determine the student's ability to "recognize words or phrases that both logically and stylistically complete the meaning of a sentence" (ETS, 1988, p. 30). The student must decide which of five word, sets of words or phrases can best complete a sentence. In completing this type of task, the student must consider which answer gives the sentence a logically satisfying meaning and stylistic integrated whole to the discourse.

Quantitative Ability (GRE-Q). The second subscore of the GRE General Test measures basic mathematical abilities, the understanding of basic mathematical concepts, the ability to reason quantitatively and to solve problems that require skills in mathematical analysis. The quantitative items seek not to exceed the abilities common to undergraduates, regardless of field of study. Questions test the student's facility with arithmetic, algebra and geometry. Questions may be in words, metric units and symbols or figures, graphs and tables.

Regular Mathematics (RM). This quantitative item-type has also been labelled Discrete Quantitative questions and Arithmetic, Algebra and Geometry in various GRE and ETS publications.

49

Quantitative Comparisons (QC). These items test the student's ability "to reason quickly and accurately about the relative sizes of two quantities or to perceive that not enough information is provided to make such a decision" (ETS, 1988, p. 34).

Data Interpretation (DI). Data interpretation items present sets of data in graphs and tables and ask students to synthesize the information, choose the correct data to answer the question, or to determine that the information needed is not present in the data set.

Analytic Ability (GRE-A). The third subscore of the GRE General Test is designed to measure students' ability to think analytically. This subscore is comprised of two item-types: Analytic Reasoning and Logical Reasoning.

Analytic Reasoning (ARE). Analytic reasoning items measure a student's ability "to understand a given structure of arbitrary relationships among fictitious persons, places, things, or events, and to deduce new information from the relationships" (ETS, 1988, p. 38).

Logical Reasoning (LR). These items assess a student's ability to understand, analyze, and evaluate positions and contentions. Specific questions may evaluate a student's ability to recognize a point of argument or the assumptions on which a position is based, to draw conclusions or form hypotheses, to assess the manner of arguments and the evidence supporting them.

While the GRE General Tests are designed to describe the student's broad verbal, mathematical, and analytical abilities, the 9 individual item-types of the Test provide discrete measures of general learned abilities. One should avoid, however, making the assumption that the GRE measures all general learned abilities associated with collegiate learning or even those intended as the educational goals of a particular college, university, program, major, or course. Nevertheless, the GRE provides a broad set of measures of general learning from which a model to assess selected gains in cognitive development of

49

college students.

Summary

In this section the purpose, scope, and method of the DCP project were described. The purpose of the research was to test the hypothesis that student enrollment in different patterns of coursework affects the development of their general learned abilities.

There was no clear consensus in the literature reviewed as to what constitutes the general learned abilities of college students. There was general agreement that the assessment of students' general cognitive development necessities multiple measures of general learn ng.

One set of such measures consists of the 9 item-types of the General Test of the Graduate Record Examination. The GRE General Test has been criticized as an assessment measure of baccalaureate learning because of the limited number and advanced abilities of the students upon which the tests were normed. Similarly, the multiple choice format of the GRE has been criticized for not measuring higher on developmental reasoning and creative thinking skills. These criticisms notwithstanding, the GRE presents a common set of measures of general collegiate learning. The strong correlation of the GRE with the Scholastic Aptitude Test (SAT) affords an opportunity to control the assessment of student learning for the comparable knowledge, skills and abilities students possessed upon admission to college. The largest amount of variance or general learned abilities logically should be attributable to learning prior to college. Nevertheless, when such learning is removed from the analysis, residual scores should provide indicators of students' development during the college years.

Just as determining what constitutes student achievement in college is problematic, so too is the determination of what the formal curriculum of a college or university is. A distinction in the DCP project is made between the

50

intended curriculum, as stated in the college catalogs and bulletins and the actual curricular record, as represented on the student transcript. A coursework pattern was defined as a set of courses whose effects on general learning are similar. This definition is an empirical artifact of student enrollment behavior, rather than an academic plan or stated curricular sequence, as might be found in a catalog or program brochure.

By clustering courses into patterns according to their effect on general learned abilities, a basis is provided for examining what students took in light of what they learned. If what the GRE measures was what the college intended as the outcomes of the general education curriculum, and if the courses shown to affect positively the general learning of undergraduates were the same as the colleges general education requirements, then that college may take pride in the evidence of the effectiveness of its curriculum. Such a comparison of what a student takes with what that individual learns is predicated upon the hypothesis that enrollment in a different pattern of coursework leads to different effects in general learning.

To test the differential coursework hypothesis, a cluster analytic model was developed. The residual scores of GRE item-types were attributed to the specific coursework of which students enrolled. Each course was described in terms of the mean of residuals of the students who had enrolled in the course. Thus, these were 9 mean item-type residuals for each course found on student transcripts. Next, courses were sorted and clustered according to these 9 criterion variables. Finally, the validity of the group was tested using discriminant analyses. From the discriminant analyses, coursework affecting general learning could be differentiated from that serving some other role (such as learning within the major or learning not measured by the GRE).

Through the development of the cluster analytic model, the effect of coursework on general learned abilities may be assessed. Furthermore, the

51

extent to which the item-type of the GRE represent discrete measures of general learning can be assessed. Similarly, the intentions of a college or university general education curriculum can be compared to empirically-derived coursework clusters found to be associated with gains in general learning. Thus, the model can be used to assess student learning to determine the strength and independence of the measures of learning selected, and to compare the intended curriculum with the actual course-taking behavior of students.

II. Methodology and Procedures:

Cluster Analytic Model

A conceptual/empirical approach was used in the selection, testing and adoption of a specific methodology for the analysis of coursework patterns. The approach was conceptual in that theoretical concepts differentiating coursework discussed in the previous section restricted the empirical approach to conform to the nature and orientation of the research problem. What follows is a discussion of the process of cluster analysis and its application to the investigation of coursework patterns; in that discussion, cluster analysis is contrasted to other statistical methods of potential value to the research investigation.

Previous transcript analysis studies have used the general linear model and regression analysis (Benbow and Stanley, 1980, 1982; Pallas and Alexander, 1983; Prather and Smith, 1976a, 1976b). The rationale for the use of regression is based upon practical and theoretical justifications. Regression analysis allows maximum design flexibility and is statistically robust. Transcript analyses involve large amounts of data. For example, Prather et al. (1976) examined 8,735 student transcripts which collectively contained 189,013 individual course grades. Regression analysis provides an effective technique for presenting the diverse nature of the data while maintaining a consistent analysis rationale. However, the general linear model does not provide a direct means of assessing the additive and temporal aspects of course patterns, as described in the previous chapter.

To distinguish cluster analysis from other approaches, certain terms need definition. The term classification is used here to refer to the categorization of the courses in which students enrolled over the duration of their baccalaure-

ate program. It is the systematic and unique way a college or university labels and arranges its courses (i.e., Honors 101, French 340, etc.); that scheme or arrangement of classes is already known in a disaggregate form on student transcripts. Identification is the allocation of individual courses to be established in categories on the basis of specific criteria (i.e., Biology 205 is classified by many universities as a sophomore level class in the department of Biology).

Discriminant analysis is a process used to differentiate between groups formed on an a priori basis (See Biglan, 1973a for an example). Discriminant analysis does not to discover groups; it identifies a set of characteristics that can significantly differentiate between the groups. The process allows the analyst to allocate new cases to one of the a priori groups with the least amount of error. In contrast, cluster analysis recovers groups representing particular patterns from diverse populations (Lorr, 1983; Romesburg, 1984). In the model developed to analyze coursework patterns, cluster analysis is used to classify courses according to student achievement criteria, while discriminant analysis is used to test and provide secondary validation of the cluster groupings and to identify those criteria which significantly differentiate one cluster of coursework from another.

Factor analysis is different from cluster analysis in that its attention is on the similarity of the variables (attributes). The aim is to identify a small number of dimensions (factors) that can account for individual differences on the various measures or attributes. Thus, the aim of factor analysis is to reduce or consolidate the number of attributes of a variable set while the purpose of a cluster analysis is simply to classify or taxonomize data into groups on the basis of a set of attributes. Miller (1969) examined 48 common nouns; through cluster analysis he identified five subgroups referring to living

things, nonliving things, quantitative terms, social interactions, and emotions. . other example of cluster analysis is Paykel's (1971) analysis of 165 depressed patients. Using symptom ratings and historical variables, he grouped the patients into four clusters: the retard psychotic, the anxious, the hostile, and the young depressive. Cluster analysis refers to a wide variety of techniques used to classify entities into homogenous subgroups on the basis of their similarities.

The end products of cluster analysis are clusters or pattern sets. Since the exact number and nature of the course patterns is not known in advance, the clustering process is actually technically preclassificatory. In other words, cluster analysis techniques are used to construct a classification scheme for unclassified data sets. In this way, cluster analysis empirically arranges the courses of a college curriculum using student decision-making behavior (as represented on transcripts) as the primary source of information. The courses are classified in a hierarchical dendrogram or tree. The relationship between courses is determined by their similarity on the criteria used in the classification. In this way, the similarity between courses is determined empirically, rather than by arbitrary concepts (i.e., "life sciences") or levels (i.e., "freshmen level survey"). This conceptual/empirical approach was selected due to the lack of agreement in the higher education literature on a common research paradigm, model or philosophy for the organization of coursework (Bergquist et al., 1981; Biglan, 1973a; Furhmann and Grasha, 1983; Gaff, 1983; Rudolph, 1977; Sloan, 1971; Veysey, 1973).

Cluster analysis conforms to the conceptual restrictions placed on the model in order to assess the effect of coursework patterns on student learning. Cluster analysis provides a statistical procedure for examining coursework using multiple criteria. It can accommodate both quantitative and qualitative attri-

butes of varying dimensions. Thus, the criterion selected need not be test scores; nominal, order, interval and ratio data have been successfully used as attributes in cluster analysis (Romesburg, 1984). Cluster analysis uses these attributes to arrive at patterns of coursework independent of any institution- ally prescribed a priori distinctions among courses. It therefore is capable of testing notions of combinations, sequence and progression of courses within the college curriculum. It leads to the discovery of clusters (or patterns) of coursework in student transcripts, based on the attributes of students' general learned abilities. Since the purpose of the cluster analytic model is to group coursework homogeneously according to its relation to student learning outcomes (Lorr, 1983; Romesburg, 1984), cluster analysis was chosen as the primary methodology for analyzing student transcripts in this Differential Coursework Patterns Project.

General procedural steps

This section describes the steps in the statistical process embodied in the cluster analytic model. These steps address the four research objectives (discussed in Section 1) of the model. Student score gains are derived. Student transcripts are examined, and courses reported on them are clustered into patterns based on the score gains of the students who enrolled in the courses. Resulting patterns of coursework are again analyzed and classified according to attributes associated with student course-selection and decision-making. Resulting patterns of coursework are analyzed and classified according to attributes associated with the educational environment of the college or university: (a) the type of college or university, as indicated by Carnegie classification (1987), (b) the type of general education degree re- quirements of the institution, as indicated in the college catalog or bulletin,

- 41 -

(c) the type of academic discipline or field of study, as indicated by the course prefix on the transcript, and (d) the student demographic character- istics, as indicated on the demographic questionnaire completed by the student at the time of GRE testing. Hypothesized patterns of coursework generated from one set of student transcripts may be validated through the replication of the cluster analytic model to a second sample of student transcripts.

Quantitative versus qualitative measures

As previously described in Section 1, there is more than one view of what constitutes representation of a college or university curriculum within a sample of student transcripts. One view suggests that only those courses in which students most frequently enroll constitutes the curriculum associated with gen- eral learning in the undergraduate program. A second view holds that any cou se offered may contribute to the general learning of students. The first view im- plies a more restricted view of the curriculum than does the second. These con- trasting views resulted in the development of two alternate procedures for as- sessing the associated effects of coursework patterns on general learned abili- ties. Reported in the following section are the results of the first procedure, the quantitative cluster analysis. The second procedure, qualitative cluster analysis, is described thereafter.

The cluster analytic model uses multiple measures of general learned abili- ties as attributes with which to classify courses taken into patterns. These attributes can be expressed quantitatively or qualitatively. For example, a sophomore level mathematics class (i.e., Math 201) can be described according to the mean residual scores of students (from the sample) who enrolled in the course. Math 201 can also be described nominally; here the researcher simply notes whether one or more students with high residual scores enrolled in the

- 42 -

course. Both the quantitative and qualitative descriptions of Math 201 serve to determine the relation of the course to other courses according to the item-type criteria variables.

When a sample of students is used to examine the effects of a particular college curriculum on general learning, there are a limited number of courses within the curriculum which can be analyzed quantitatively. The GSU Historical Group example, previously described, illustrated this problem. Only a limited percent of all courses appearing on the sample transcripts can be analyzed if the number of students enrolling in a given course is a concern in the analysis. However, such quantitative analysis of the curriculum can yield much more accurate information regarding the effect a particular course may have on a given measure of student general learned ability. To generalize about a course on the basis of 5 or more student residual scores provides a level of information that far exceeds that of simply noting whether any student who performed well on a given measure enrolled in that course.

There are advantages and disadvantages to either the quantitative or the qualitative approach. In the quantitative analysis, a limited number of courses can be examined, but, in practice, those courses are the ones in which most students enroll and encompass all those in which students are required to enroll. Math 101, a required mathematics course in a college's curriculum, would be included in those courses examined in a quantitative cluster analysis since all students are required to enroll, while Math 450 designed primarily for senior level math majors would not be included -- assuming the sample of students is random and not confined to mathematics students.

There are those, however, who may argue that it is the advanced coursework within a given discipline which facilitates general student learning. It has been suggested that the study of liberal arts disciplines teaches students a

- 43 -

mode of inquiry which facilitates their learning of other forms of knowledge, abilities and skills (Biglan, 1973a, 1973b). Similarly, courses with traditionally restricted enrollments may not appear in an analysis of coursework selected by the frequency of enrollment. Analysis of the effect of credit for study abroad or honors programs or the assessment of coursework patterns of specific groups of students might not be possible. Therefore, under these and related circumstances, it is desirable also to examine as many courses of a student's transcript as possible, rather than restricting the analysis to only those courses in which students most frequently enroll.

Examination of all courses on a student's transcript may not be feasible. Some courses may have only one student enrolled from the sample group, the cohort, or population of students examined. Recall that in the cluster analytic model, a student's GRE item-type residuals are attributed to all the courses in which he/she enrolled. The contribution of individual courses to the curriculum is calculated as the sum of the effects of the students who enrolled in those courses. Courses with low enrollments from the sample group or the group being examined have higher margins of error because the effects are discerned from a smaller number of students. Thus, courses with an enrollment of one student from the sample group do not provide a basis for quantitative analysis, while courses with limited enrollment (2 or more) may be amenable to the treatment of that enrollment solely as a nominal variable.

In a quantitative cluster analysis, the metrics used for each course are the mean GRE item-type residuals which contain interval information about the gains of students who enrolled in the course. In a qualitative cluster analysis, the metrics used are whether students with high score residuals did or did not enroll; the metric is reduced to a dichotomous nominal variable. There is a trade-off in a qualitative cluster analysis between inclusiveness of the

curriculum and precision of the information.

Any quantitative attribute, such as a GRE item-type residual, can be dichotomized and converted into a binary attribute (Anderberg, 1973). Such a procedure lessens the precision of information in the data set because the process is irreversible. The data from an interval scale is collapsed into a nominal one. It is commonly held that ratio scales provide more precise information than interval scales, that interval scales are more precise than ordinal ones, and that all the preceding are more informative than nominal scales. However, the choice of scales is constrained by different factors.

First, institutional researchers are often under monetary constraints. The costs of obtaining test scores for all college graduates, for example, may not be feasible on an on-going basis. Hence, it may not be practical to gather the number of student transcripts and assessment information needed to use the quantitative cluster analysis with courses other than those in which students most frequently enroll.

Second, institutional researchers have a choice between an intensively detailed picture of the curriculum using the ratio data of mean residuals or a less detailed picture provided by binary information. As has been previously discussed, there are occasions when the scope of the analysis is to be preferred over the precision of the analysis.

Third, "data do not automatically inform the researcher" (Romesburg, 1984). To have meaning, transcript and test data must be interpretable within a curricular context. The primary question is, "Which coursework patterns contribute to general student learning?" The secondary questions are, "How much do the patterns contribute?" and "What is their relative contribution?" Qualitative analyses are not categorically inferior. In this case, a qualitative analytic question precedes the one which may be answered

quantitatively.

Procedure 1: Quantitative cluster analysis

Described below are steps required in Procedure 1 (Quantitative Analysis) to assess the effects associated with the coursework patterns on the general learned abilities of college students. The research design uses as data sources transcripts and GRE and SAT test scores from a sample of students. The 9 item-type categories of the General Tests of the Graduate Record Examination are used as measures of general learned abilities of college seniors. These seniors' SAT scores are used as variables to control for the academic abilities of these students when they first entered college. The student transcripts are used as the record of the sequence of programs in which these seniors enrolled.

The first objective of the cluster analytic model is to determine the student gains in general learned abilities over the time of their baccalaureate program. To do this, first the residual score of each item-type for each student is calculated; the residual score is the difference between the student's actual score and the score predicted by the student's corresponding SAT score. Thus, for each student outcome measure there is a student score gain for each person in the sample group.

The second objective is to determine patterns of coursework on the student transcripts which are associated with student score gains. This is accomplished through cluster analysis, using student score gains (GRE item-type residuals) as attributes of the programs in which students enrolled.

A raw data matrix consisting of columns of programs and rows of score gains is created. The mean residual score for all the students in the sample who enrolled in a given program is calculated and becomes the metric value for that program. The correlation coefficient is used as the resemblance coefficient to

- 46 -

transform the data matrix into a resemblance matrix, wherein the similarity of

residual scores for students enrolling in one program can be compared with those

enrolled in another program. Once the resemblance matrix indicating the propor-

tional relationship of programs is established, a clustering method is selected

and executed to arrange a tree or dendrogram of programs related by the student

score gains. Next, a discriminant analysis is performed on the resulting

clusters of coursework to: (a) determine the extent to which the programs have

been correctly classified according to the 9 mean student residual scores, (b)

to determine which of the 9 mean residual scores were correlated with particular

discriminant functions, and (c) to determine which coursework clusters exhibited

high mean residual scores relative to each discriminant function. From the

discriminant analysis an association can be inferred between coursework patterns

(clusters) and general learned abilities (student score gains on 9 criterion

variables). The cluster-analytic procedure groups programs frequently chosen by

students according to the strength of their associated effect on the student

score gains.

Described in greater detail below the steps followed in this cluster ana-

lytic procedure:

Step 1. Calculate a student residual score for each item-type (attribute) of

each student GRE. This step removes the predictive effect of the

student's SAT scores from the GRE item-type, thereby controlling for the

academic ability of the student upon entrance to college. For GRE

Quantitative item-types, the effect of the student's SAT Math score is

partialled out. For the GRE Verbal item-types, the effects of the SAT

Verbal score is partialled out. For the GRE Analytic item-types, the

effect of the combined SAT Verbal and SAT Math scores are partialled

- 47 -

62

out. In this way, the student's academic abilities prior to entering college is controlled when calculating student residual scores.

Step 2. Calculate the mean residual score for each program enrolling 5 or more students from the sample group. Cross-listed programs are standardized so that they have only one identifier. Cross-listed programs include those with identical numbers that have different labels. Programs with the same program identifier but with abstentiously different content (i.e., "Music 101: Voice" and "Music 101: Piano") are excluded from the analysis. However, catalog changes are accounted for. If Math 201 in 1982 was renumbered as Math 211 in 1985, Math 201 and Math 211 for those years are treated as the same program for the purposes of analysis.

The proportion of programs included in the analysis is related to a) the extensiveness of the program listings in the curriculum, and b) the size of the student sample. The more extensive the curriculum, the less frequently 5 or more students from the sample will have enrolled in the same program. Likewise, the smaller the size of the student sample, the less frequently 5 or more students from the sample will have enrolled in the same program.

Step 3. Create a raw data matrix by using the mean residual scores for the programs found on 5 or more of the student transcripts. The rows in the data matrix consist of the 9 GRE item-type scores while the columns represent those programs enrolling 5 or more students. Each cell value of the matrix is a mean GRE item-type residual score for those sample group students enrolling in a specific program. For example, the

program (object) in the first column in the data matrix is ANTHROPOLOGY 101, and the student outcome measure in the first row of the data matrix is DATA INTERPRETATION. The student score gains are .40, .45, .50, .55, and .60; the mean score gain, therefore, is .50 and is entered as the metric variable in cell (1,1) of the matrix. Since the variables in each row are of the same magnitude, and therefore, have comparable effect on the resulting cluster analysis, the data matrix does not need to be standardized (Romesburg, 1984). The cluster analysis will taxonomize programs in the curriculum according to whether students who showed positive residuals on each item-type were enrolled in the programs. This step prepares a raw data matrix to be used in a general cluster analysis based on quantitative data.

Step 4. Select a resemblance coefficient. The resemblance coefficient (Romesburg, 1984) is also called the similarity index (Lorr, 1983). The purpose of the resemblance coefficient is to explain the similarity (or dissimilarity) of each cell to each of the other cells in the data matrix; it is expressed mathematically. There are many resemblance coefficients; each will express the similarity between programs (objects) in a slightly differently way. Each coefficient is appropriate for achieving slightly different research goals.

The resemblance coefficient selected for this study is Pearson's product-moment correlation coefficient. It is appropriate for use with ratio data. The resemblance coefficient indicates the similarity of programs to each other according to the 9 item-type residuals (attributes) coded in the data matrix. The resemblance coefficient

- 49 -   64

expresses the relationship of two programs proportionally.

Step 5. Calculate a resemblance matrix from the raw data matrix. The resemblance matrix is calculated by transforming the raw data matrix using the correlation resemblance coefficient. In the cluster analytic model, the data matrix consists of quantitative data described by 9 attributes ranging in value from 1.00 to -1.00. In the resemblance matrix, the columns represent the first program (object) in a pair, the rows represent the second program (object) in a pair. The resemblance coefficient (Pearson's r) is entered into each cell. The cell value represents the extent to which the attributes on the first program explain the variance in attributes on the second program. The resemblance coefficient serves as a measure of similarity between one program and each other program in the calculation of clusters or coursework patterns.

Step 6. Select and execute the clustering method. A resemblance matrix is transformed into a tree of related programs (objects) by use of a clustering method which is a series of steps that removes values from the resemblance matrix. Therefore the size of the matrix is reduced. Each time a value is removed from the resemblance matrix it is placed in the cluster tree or dendrogram. In the last step, the resemblance matrix disappears completely and the tree is completed as the last value is inserted.

Romesburg (1984, p. 139) recommends the unweighted pair-group method using arithmetic averages (UPGMA), also known as the average linkage

method. UPGMA is recommended over single linkage clustering method
(SLINK) and complete linkage clustering method (CLINK) for two reasons.
First, it can be used with any resemblance coefficient, while SLINK and
CLINK are designed to be used with interval and ratio data in a
quantitative data matrix. Second, it judges the similarity between
pairs of clusters in a less extreme manner than do SLINK and CLINK. The
average linkage method (UPGMA) is available on SPSSx, SAS and BMDP
statistical packages.

Step 7. Determine the optimum number of coursework clusters. Cluster analysis
is a procedure for taxonomizing or classifying coursework data. The
number of groups or patterns in which the data is classified according
to the criterion variables is an arbitrary one. Once relationships bet-
ween programs have been determined, the researcher must decide on how
many groups in which to put the data. Discriminant analysis provides a
means to test the secondary validity of the coursework pattern
groupings.

By computing successive cluster analyses for different numbers of clus-
ters and then conducting discriminant analyses on the resultant group-
ings, one can identify the number of clusters which has the highest pre-
dictive value, given the criterion variables used. Using the DISCRIMIN-
ANT program in SPSSx, for example, will identify how many members of
each coursework pattern or cluster were correctly classified, how many
could be classified in other patterns, and what was the overall percen-
tage of correct classification.

- 86

The number of clusters with the highest predictive value may not be the sole objective in examining the merits of different cluster solutions to the cluster analysis. Theoretically, a four cluster solution may have high predictive value for GRE item-types because the item-type residuals are forced into three discriminant functions which should approximate the GRE sub-scores. Likewise, a 10 cluster solution may prove to be slightly less predictive, but the 9 GRE item-type residuals may be more clearly associated with discrete coursework patterns. Careful visual inspection of the cluster dendrogram often suggests appropriate cluster solutions to test using discriminant analysis.

Step 8. Determine which criterion variables contribute significantly to which discriminant functions. DISCRIMINANT in SPSSx, for example, calculates the pooled within-groups correlations between the discriminating variables (in this case, the mean residual scores on the 9 item-types) and the canonical discriminate functions. Large positive and negative correlations are identified. Eigenvalues for each discriminant function are assessed. Eigenvalues express the proportion of variance in student scores explained by the discriminant function. Discriminant functions that explain less than 5 percent of residual score variance or that have a probability of error exceeding .001 are discarded. Next, the group means for each coursework cluster can be examined. In this manner, the patterns of coursework associated with one or more mean item-type residual scores can be identified.

<u>Step 9</u>. Repeat Steps 1 to 8 using a second cohort of students. Following Steps
1 through 8 will produce a set of <u>hypothesized</u> relationships between
coursework patterns and student score gains on 9 criterion measures of
general learned abilities. Hypothesized relationships cannot be tested
or validated using the same data. Therefore, a second institutional
sample is drawn. A second group of students are tested and a second set
of transcripts and student score gains are evaluated. Repeated use of
the model should refine and clarify members within each coursework
pattern.

Through the above 9 steps, the cluster analytic model classifies the most
frequently enrolled programs according to their associated effect on student
score residuals. Procedure 1 classifies programs according to a ratio index of
similarity to other programs.

Procedure 2: Qualitative cluster analysis

While the Quantitative Cluster Analysis Procedure (Procedure 1) examined
those programs enrolling 5 or more students, it may be desirable, and in this
case necessary, to associate a greater proportion of the curriculum with the
students who showed significant improvement on one or more of the GRE
item-types. A second Cluster Analytic Procedure was developed. Procedure 2 was
labeled "Qualitative Cluster Analysis". In Procedure 2 only, programs enrolling
2 or more students with positive residuals were considered.

Due to the small number of students in the Evergreen State Samples,
Procedure 1 was not utilized. The qualitative cluster analytic procedure
described in preceding progress reports (Ratcliff, 1988a, 1988b) dichotomized
the student sample into subgroups of students: those scoring at or above the

- 53 -

mean on a given attribute and those scoring below the mean. Separate cluster analyses were then conducted on each sub-group and the results of the two were then compared. It was discovered that such a procedure tended to exclude the programs with the highest enrollment levels, since the higher the enrollment, the greater the probability that students from both the high-residuals group and the low-scoring residuals group enrolled in the program. Conversely, in low enrollment classes, the probability that students from both groups enrolled was markedly reduced. For this reason, further testing and development of the qualitative cluster analytic procedure was warranted.

The steps required in Procedure 2 to classify coursework patterns based on student residual scores in general learned abilities are described below. As in Procedure 1, student GRE item-type residuals first are computed for each student outcome measure. Second, those students who scored at or above the mean of student residuals for a given item-type are identified. Third, the proportion of students with residuals at or above the mean relative to the total student sample is computed for each item-type for each program. The standard error of estimate for that proportion is calculated. The metric for each attribute (item-type) for each program thus becomes the proportion of students enrolling in the program whose residuals are at or above the mean. In cases where the standard error is zero, a constant--empirically determined-is used to avoid the mathematical problem of dividing by zero.

Once the metric of the proportion of high-residual students enrolled in a given program is calculated for each item-type, a raw data matrix consisting of columns of programs and rows of proportions is created. The proportion of high-residual students in the sample becomes the metric value for that program. The correlation coefficient is used as the resemblance coefficient to transform the data matrix into a resemblance matrix. In the resemblance matrix the

- 54 -
69

similarity of proportions of high-residual students enrolling in one program can be compared with those enrolled in another program. Once the resemblance matrix of programs is established, then a clustering method is selected and executed to arrange a tree or dendrogram of programs related by the student residual scores. Next, a discriminant analysis is performed on the resulting clusters of coursework to: (a) determine the extent to which the programs have been correctly classified according to the 9 mean student residual scores, (b) to determine which of the 9 mean residual scores were correlated with particular discriminant functions, and (c) to determine which coursework clusters exhibited high mean residual scores relative to each discriminant function. From the discriminant analysis, an association between coursework patterns (clusters) and general learned abilities (student residual scores on 9 criterion variables) can be inferred. The qualitative cluster-analytic procedure groups programs according to the weighted proportion of students with high residuals. Thus, programs with larger proportions of students demonstrating at or above average gains are classified together.

Below are the steps in Procedure 2 of the cluster analytic model:

Step 1. Calculate a residual score for each item-type (attribute) of each student GRE. This step is identical to that of Procedure 1 and removes the predictive effect of the student's SAT scores from the GRE item-type, thereby controlling for the academic ability of the student upon entrance to college.

Step 2. Compile the residual score of each student at or above the mean for each item-type (attribute). Calculate the proportion of students at or above the mean enrolling in each program and the standard

error of estimate for that proportion. Weight the proportion of students at or above the mean for a given program and a given item-type by the reciprocal of the standard error of estimate for that proportion. Exclude programs with an enrollment of 1 or no students, since a standard deviation cannot be calculated from an enrollment of one.

Step 3. Construct a data matrix with columns consisting of all the programs (objects) with an enrollment of 2 or more students, as evidenced by the transcripts of the students in the sample. The rows denoted each of the 9 item-type residuals. The cells in the data matrix consist of the weighted proportion of students whose residual scores were at or above the mean for the given item-type (attribute). Cross-listed programs are standardized so that they have only one identifier.

Step 4. Select a resemblance coefficient. The same resemblance coefficient used in Procedure 1 can be used in Procedure 2: Pearson's product-moment correlation coefficient. This will indicate the similarity of programs to each other according to the weighted proportion of students at or above the mean on each of the 9 item-type residuals (attributes), as coded in the data matrix.

Step 5. Calculate a resemblance matrix from the raw data matrix. As in Procedure 1, the resemblance matrix is calculated by transforming the raw data matrix using the correlation resemblance coefficient. The cell value represents the extent to which the weighted proportion of students

- 56 -

71

at or above the mean on the attributes in the first program explain the variance in the weighted proportion of students at or above the mean on the attributes of the second program. Thus, the resemblance coefficient serves a measure of similarity between one program and each other program in the calculation of clusters or coursework patterns.

Step 6. Select and execute the clustering method. A resemblance matrix is transformed into a dendrogram or tree of related programs (objects) by the use of a clustering method which is a series of steps that removes values from the resemblance matrix. Therefore, the size of the matrix is reduced. Each time a value is removed from the resemblance matrix it is placed in the tree. In the last step, the resemblance matrix disappears completely and the tree is completed as the last value is inserted.

As in Procedure 1, the unweighted pair-group method using arith-metic averages (UPGMA), also known as the average linkage method, is used as the clustering method. Recall, UPGMA is available on SAS, SPSSx, and BMDP statistical packages.

Step 7. Determine the optimum number of coursework clusters. As in Procedure 1, discriminant analysis is used to provide a means to test the secondary validity of the coursework pattern groupings.

Also as in Procedure 1, the number of clusters with the highest predic-tive value may not be the sole objective in examining the merits of dif-ferent cluster solutions to the cluster analysis. Cluster groupings

should be sought which most clearly disclose the relationship between
the criterion variables and the coursework patterns contained in the
cluster tree. Careful visual inspection of the cluster dendrogram often
suggests appropriate cluster solutions to test using discriminant analysis.

Step 8. Determine which criterion variables contribute significantly to which
discriminant functions. This step is identical to that in Procedure 1.

Step 9. As in Procedure 1, repeat Steps 1 to 8 using a second cohort of students
to determine if the patterns found in the first cohort can be replicated
in the second.

Through the above 9 steps of Procedure 2, the cluster analytic model classifies all programs with an enrollment of 2 or more students (from a sample of
transcripts) according to the proportion of students enrolling who evidenced
gains at or above the mean on the selected measures of general learned ability.

Procedure 2 allows for the examination of a greater proportion of the curriculum than Procedure 1, but in doing so, it reduces the precision of information used in the analysis through the transformation of quantitative data on
student residuals into a qualitative dichotomy (between those students whose
residuals were at or above the mean and those whose residuals were not). Due to
the small number of students in the Evergreen Samples, Procedure 1 was not
used. Instead Procedure 2 was utilized to analyze the sample transcripts.

Describing the resulting coursework patterns

The third objective of the cluster analytic model is to describe the patterns of programs resulting from Procedure 2 analysis according to: sequences and combinations of programs within the cluster, according to term of enrollment data found on student transcripts. Can the program patterns resulting from the Procedure 2 analysis be meaningfully described by the above factors? Every program in each resulting pattern can be described according to any of the above factors, just as it was described by associated student residual scores. The extent to which programs in one cluster differ from other clusters on any of the factors can then be tested using analysis of variance (ANOVA) or the Kruskal-Wallis one-way analysis of variance for ordinal data.

III. Description of the Evergreen State College Combined Samples

This section describes the Evergreen State combined samples (1987-88, 1988-89) group of graduating seniors. The criteria for inclusion in the samples was a graduation date or estimated graduation date during the 1987-1988 academic year for the first sample. The same criteria, only one year later, was used for the second sample. There were no transfer students or courses included in the analysis.

Combined Samples #1 and #2

Evergreen State combined samples consisted of the transcripts and test scores for 20 students from sample #1 and 26 students from sample #2.

Gender has been shown to be a significant factor in the academic performance of college undergraduates. Over one-half (54.3%) of the sample were female, while 43.5 percent of the population were male (see Figure 3-1).

Race and ethnicity also have been shown to be strong predictors of academic performance. Seventy-six point one percent of the combined Sample were white, 2.2 percent were Asian, and 2.2 percent described themselves as "other". Nineteen point six percent declined to describe their ethnicity (see Figure 3-2).

Major field of study has been shown to be correlated to performance in the GRE examinations. Many disciplines dominated the self-reported majors for the GRE background questionnaire with 41.3% reporting interdisciplinary programs. Majors in Computer Science, Pre-Med, Drama, Humanities, and Radio, TV, and Film also were reported by more than one student (see Figure 3-3).

Figure 3-4 presents the SAT scores for the combined sample. The average SAT Verbal for the combined sample was 529, compared with an average SAT Math

score of 513, indicating a sample that appears better prepared in verbal skills than math skills.

Figure 3-1. Distribution of Evergreen State Combined Sample: Gender
==

| | Combined Sample | |
Gender	N	PERCENT
Female	25	54.3%
Male	20	43.5%
NO RESPONSE	1	
TOTALS	46	100.0%

==

Figure 3-2. Distribution of Evergreen State Combined Sample: Ethnicity
==

| | Combined Sample | |
Ethnicity	N	PERCENT
Not specified	9	19.57%
Black	0	.00%
Native American	0	.00%
Asian	1	2.17%
Hispanic	0	.00%
White/Non-Hispanic	35	76.09%
Foreign	0	.00%
Other	1	2.17%
TOTALS	46	100.00%

==

76

Figure 3-3. Distribution--Evergreen State Combined Samples: First Major

Major	Evergreen Combined Samples	
	N	PERCENT
No Response	5	10.9%
Biology	1	2.2%
Computer Science	2	4.3%
Pre-Med	2	4.3%
Systems Engineering	1	2.2%
Art	1	2.2%
Drama/Theater Arts	2	4.3%
Arts-Performance/Studio	1	2.2%
Creative Writing	1	2.2%
American History	1	2.2%
History-Other	1	2.2%
Humanities and Arts-Other	2	4.3%
Elementary-level Teaching	1	2.2%
Secondary-level Teaching	1	2.2%
Business Management-Other	1	2.2%
Radio, TV, and Film	2	4.3%
Communications-Other	1	2.2%
Interdisciplinary Programs	19	41.3%
Field Not Classified Above	1	2.2%
TOTALS	46	100.0%

Figure 3-4. Summary of SAT Scores for Evergreen Combined Samples

SAT Part Score	N	Mean	Standard Deviation	Range	Standard Error
Verbal	46	528.91	93.29	360-730	13.75
Math	46	512.61	84.47	330-670	12.46
SAT Total	46	1041.52	157.35	750-1330 .	23.20

SAT Verbal Scores Midpoints	N	Percent	Distribution
350	2	4.8%	****
400	3	7.1%	******
450	10	23.8%	*********************
500	8	19.0%	*****************
550	10	23.8%	*********************
600	6	14.3%	************
650	3	7.1%	******
700	3	7.1%	******
750	1	2.4%	**

SAT Math Score Midpoints	N	Percent	Distribution
350	2	4.3%	****
400	5	10.9%	*********
450	9	19.6%	*****************
500	10	21.7%	******************
550	9	19.6%	*****************
600	5	10.9%	*********
655	6	13.0%	***********
700	0	.0%	

Figure 3-5. Entering Semester of Evergreen Combined Sample
==

Entering Semester	N	PERCENT
Fall 1980	1	2.2%
Fall 1981	1	2.2%
Fall 1982	2	4.3%
Fall 1983	3	6.5%
Summ 1984	1	2.2%
Fall 1984	16	34.8%
Spr 1985	1	2.2%
Fall 1985	20	43.5%
Fall 1986	1	2.2%
TOTALS	46	100.0%

==

Figure 3-6. Planned Year of Graduation: Evergreen Combined Sample
==

Planned Year of Graduation	N	PERCENT
No Response	2	4.3%
1987	1	2.2%
1988	20	43.5%
1989	22	47.8%
1990	1	2.2%
TOTALS	46	100.0%

==

Figure 3-5 shows that the overwhelming majority of the combined sample entered the institution in the fall of 1984 or the fall of 1985. All but one student in the combined sample were projected to meet graduation requirements for a bachelor's degree either during the 1987-88 (sample #1) or the 1988-89 academic year (Figure 3-6). The one student gave a self projection of graduation in 1990.

Students in the combined sample were clearly planning some form of post-baccalaureate study (Figure 3-7). One-half planned to pursue a master's degree, while nearly one-quarter (21.7%) planned to enter a doctoral program. Only 6.5 percent had no plans for subsequent graduate study.

The educational attainment of parents has been shown to be positively correlated to student achievement in college. Only one-tenth (10.9%) of the fathers and two percent of the mothers of students had not attained a high school diploma. Nearly seven-tenths (69.6%) of fathers and one-half of mothers had attained at least an associate degree (Figure 3-8).

Figure 3-7. Degree Objectives for Evergreen Combined Sample

Degree Objectives	N	PERCENT
No Response	3	6.5%
Non-degree Study	4	8.7%
Master's Degree	23	50.0%
Intermediate Degree (e.g., Specialist)	1	2.2%
Doctorate (Ph.D., Ed.D.)	10	21.7%
Postdoctoral Study	5	10.9%
TOTALS	46	100.0%

Figure 3-8. Educational Attainment of Parents for Evergreen Combined Sample

Highest Level of Education Completed	Father N	Father Percent	Mother N	Mother Percent
No Response	1	2.2%	1	2.2%
Grade School or Less	1	2.2%	0	.0%
Some High School	3	6.5%	0	.0%
High School Diploma or Equivalent	3	6.5%	5	10.9%
Business or Trade School	1	2.2%	2	4.3%
Some College	5	10.9%	12	26.1%
Associate Degree	3	6.5%	3	6.5%
Bachelor's Degree	5	10.9%	9	19.6%
Some Graduate/Professional School	6	13.0%	3	6.5%
Graduate/Professional Degree	18	39.1%	11	23.9%
TOTALS	46	100.0%	46	100.0%

81

Figure 3-9. Prior Full-time Work Experience of Evergreen Combined Sample

Years of Work/Military Experience	N	Percent
No response	1	2.2%
None	7	15.2%
Summer(s)	21	45.7%
Less than one year	5	10.9%
One year or more, but less than 3	8	17.4%
More than 3 years, but less than 5	4	8.7%
More than 5 years, but less than 7	0	.0%
More than 7 years	0	.0%
TOTALS	46	100.0%

Figure 3-10. Extent of Community Service Activities for Combined Sample

Hours/Week in Community Service	N	PERCENT
No response	1	2.2%
0 hours	21	45.7%
1-5 hours	1b	39.1%
6-10 hours	4	8.7%
11-20 hours	2	4.3%
More than 20 hours	0	.0%
TOTALS	46	100.0%

Figure 3-11. Important Honors and Awards for Evergreen Combined Sample
==

Type of Honor/Award	N	Percent
No response	2	4.3%
Student government or organization	2	4.3%
Professional (an award or prize for field work or publication of a scholarly article or book)	0	.0%
Community service (election or appointment to a community service unit, activity, or group)	4	8.7%
Literary (editing the college paper, yearbook, or literary magazine or having a poem, story, or article published in a public paper or magazine)	5	10.9%
Artistic (a high rating in a music contest, a part in a play, opera, or show, or an award in an art competition)	11	23.9%
Scientific	2	4.3%
Athletics (a letter in athletics)	4	8.7%
None of the above categories	16	34.8%
TOTALS	46	100.0%

==

As Figure 3-9 shows, students in the combined sample possessed some full-time work experience. Over one-third (37.0%) of these students had held jobs that were more than summer jobs, while only seventeen point four had no full time work experience. Over one-half (52.2%) of the combined sample students had performed some community service during the past year, but for the majority of these students this had comprised less than five hours per week (see Figure 3-10).

Nearly two-thirds (60.9%) of the sample students had earned some form of scientific, professional, community service, literary, artistic, athletics, or

83

student government honor, or award. Prior research had shown such distinctions
to be highly correlated to student performance, persistence, progress, and
degree attainment in college (see Figure 3-11).

```
============================================================================
Figure 3-12. Summary of Evergreen Combined Sample
----------------------------------------------------------------------------
Sample Size:    46 students

       Sex:     25 females (54.3%) and 20 males (43.5%)

      Race:     35 out of 46 are white (76.1%)

Major Area:     19 of 46 students reported Interdisciplinay Programs
                as their major.  Other majors included Computer Science,
                Pre-Med, Drama, and Radio, TV, and Film.

============================================================================
```

Figure 3-12 summarizes the characteristics of Evergreen State College
combined sample group. These students were largely analogous to the population
with regard to major and race. There was an over-representation of women
(10.4%) and some variation in the representation of majors in the Sample. The
precollege achievement variables also showed some discrepency from the
population.

IV. Determining Student Learned Abilities

GRE residual scores

To control for the effects of the incoming ability of students, the predictive effect of SAT scores were partialled from GRE item-type scores. For this, 9 GRE item-type residual scores were developed as follows:

GRE Verbal item-type residuals;
    ANA: Analogies   18 questions
      SC: Sentence Completion   14 questions
      RD: Reading Comprehension 22 questions
    ANT: Antonyms   22 questions

GRE Quantitative item-type residuals;
    QC: Quantitative Comparison 30 questions
    RM: Regular Mathematics   17 questions
    DI: Data Interpretation   10 questions

GRE Analytical item-type residuals;
    ARE: Analytical Reasoning   38 questions
    LR: Logical Reasoning   12 questions

Each of the 4 GRE Verbal item-type scores were regressed on the SAT Verbal scores. Each of the 3 GRE quantitative item-type scores were regressed on the SAT mathematics scores. Each of the 2 GRE analytical item-type scores were regressed on the SAT total scores. These GRE item-type residual scores were referred to as student residual scores, that is, the improvement students showed in general learned abilities from the time they entered college to the time of GRE testing during their senior year.

Reliability and correlation of GRE item-types

Prior to partialling the effects of the students' SAT scores from their GRE item-type scores, the reliability of the GPE item-types for this sample was tested. Next, the correlation between the GRE item-types and the SAT sub-scores and total score was examined. Finally, a regression of GRE item-types on SAT

sub-scores was conducted to calculate student residual scores for each GRE item-type.

A preliminary question in the analysis of GRE item-types and sub-tests is their reliability within the sample group. Three factors typically contribute to the reliability or unreliability of test scores (Ebel, 1972). The first factor is the appropriateness and definitiveness of the questions. On one hand, the appropriateness of the questions is presumed by the widespread acceptance of the GRE as an examination used in graduate school admissions. On the other hand, the appropriateness of the items and item-types may be questioned relative to the goals of the general education curriculum of the institution. In this sense, the reliability of the GRE may vary from institution to institution.

A second factor contributing to the reliability of test scores is the consistency and objectivity of the person (or in this case, machine) who scores the examinations. All the test responses are read by an optimal scanner and scored by a computer at Educational Testing Service. The accuracy of this equipment relative to the task was presumed and not tested. According to procedures established by the Educational Testing Service for the administration of the Graduate Record Examination, all students were tested on the first Saturday morning (8:00 a.m. to 12:00 noon) in February of each test year.

A third factor contributing to the reliability is the constancy or stability of a student's ability to perform the tasks presented in the test. Students may vary from hour to hour or from day to day in their alertness, energy and recall; these may affect test performance, reducing the reliability of the scores.

Reliability is not merely the property of the GRE itself but rather of the individual item-types relative to the student group examined. The more appropriate the test is to the group of students, the higher the reliability of

- 71 -

86

the scores. Ideally, the reliablity of a set of scores may be determined using the correlation coefficient between that set of scores and another set from an equivalent test of the members of the same group. In many testing situations, including the ones described in this report, a test-retest method of determining the reliability of GRE item-types was not available.

The Guttman Split-Half method of determining reliability estimates reliability by splitting the sample into halves and determining the correlation between the scores in the two groups. The results of the split-half method are dependent upon the manner in which the group is halved. Cronbach's alpha is a statistic designed to overcome this problem. It is a generalized formula representing the average correlation obtained from all possible split-half reliability estimates.

The results of the reliability analysis for Evergreen State Sample #1 is presented in Figure 4-1a; the reliability analysis for Evergreen State Sanuple #2 is displayed in Figure 4-1b. Since different forms of the GRE were used each year, the reliability results are presented for each individual sample. For the purposes of this study, reliability coefficients at or above $a= .65$ were deemed satisfactory (Mehrens & Lehmann, 1969). Due to the exploratory nature of this research, lower reliability coefficients were accepted. In Sample Group #1, Analogies, Reading Comprehension, Data Interpretation, and Logical Reasoning evidenced low reliability. In Sample Group #2, Data Interpretation, Regular Mathematics, and Logical Reasoning showed low reliability. In both samples, the reliability of the individual item-types tended to increase with the number of items comprising the given item-type. In the cluster analytic model, the SAT sub-scores are used as measures of entering student ability. Prior to regressing GRE item-type scores on SAT scores, it is important to determine the extent to which GRE item-types and SAT sub-scores are correlated. For example,

determining whether the GRE item-type, Analogies, has a stronger correlation with SAT Verbal, SAT Math or the total SAT scores will help determine which SAT score should be used in the subsequent regression analysis.

Figure 4-1a. Reliability of Coefficients of GRE Item-Types--Sample #1

GRE Item-types	Code	Number of items	Cronbach's Alpha	Guttman's Split-half
Analogy	ANA	18	-.0087	-.4342
Sentence Completion	SC	14	.6139	.5106
Reading Comprehension	RD	22	.5533	.3407
Antonyms	ANT	22	.7514	.6489
Quantitative Comparison	QC	30	.6700	.6213
Regular Mathematics	RM	20	.7471	.8579
Data Interpretation	DI	10	.3546	.3022
Analytical Reasoning	ARE	38	.7257	.7509
Logical Reasoning	LR	12	.5419	.5247
GRE Verbal	GRE-V	76	.8093	.7129
GRE Quantitative	GRE-Q	60	.8311	.9001
GRE Analytic	GRE-A	50	.7921	.8621

Figure 4-1b. Reliability of Coefficients of GRE Item-Types--Sample #2

GRE Item-types	Code	Number of items	Cronbach's Alpha	Guttman's Split-half
Analogy	ANA	18	.6427	.6664
Sentence Completion	SC	14	.6780	.7205
Reading Comprehension	RD	22	.6080	.5133
Antonyms	ANT	22	.6432	.6794
Quantitative Comparison	QC	30	.7745	.6041
Regular Mathematics	RM	20	.6322	.6768
Data Interpretation	DI	10	.5164	.5197
Analytical Reasoning	ARE	38	.7795	.7958
Logical Reasoning	LR	12	.5388	.2754
GRE Verbal	GRE-V	76	.8446	.8398
GRE Quantitative	GRE-Q	60	.8633	.7833
GRE Analytic	GRE-A	50	.8071	.7740

88

Figure 4-2 indicates strong, positive relationships between GRE item-types and SAT scores. For the Evergreen State Combined Sample, GRE Verbal item-types were strongly correlated to the SAT Verbal sub-score with r ranging from .58 to .72. GRE Quantitative item-types had strong correlations with the SAT Mathematics sub-score, r ranging from .40 to .81. GRE Analytic item-types evidenced strong correlations with the SAT Total score (r = .66 and .71). Results of the correlation analysis were comparable to those found in the other institutional samples previously analyzed (Clayton State, Georgia State, Ithaca, Mills, and Stanford Sample Groups).

Figure 4-2. Correlation of GRE Item-Types & SAT Scores-Evergreen Combined
===

GRE Item-types	Code	SAT Verbal		SAT Math		SAT Total	
Analogy	ANA	.6747	▲	.4898		.6630	
Sentence Completion	SC	.6707	▲	.3709		.5967	
Reading Comprehension	RD	.5807	▲	.5346		.6313	
Antonyms	ANT	.7230	▲	.5102		.7026	
Quantitative Comparison	QC	.4824		.8085	▲	.7200	
Regular Mathematics	RM	.2374		.6088	▲	.4676	
Data Interpretation	DI	.1289		.3978	²	.2900	
Analytical Reasoning	ARE	.5981		.6594		.7086	▲
Logical Reasoning	LR	.5934		.5822		.6643	▲
GRE Verbal	GRE-V	.8479	▲	.6212	²	.8362	▲
GRE Quantitative	GRE-Q	.3919	²	.7761	▲	.6489	▲
GRE Analytic	GRE-A	.6545	▲	.6986	▲	.7630	▲
Minimum		.1289		.3709		.2900	
Maximum		.7230		.8085		.7200	
Mean		.5210		.5514		.6049	

[1]p < .05   [3]p < .001
[2]p < .01   [4]p < .0001
===

89

Intercorrelation of GRE item-types

The internal validity of GRE item-types can be measured by comparing the intercorrelation coefficients of GRE item-types. In the Evergreen State Sample the intercorrelations between GRE Quantitative item-types were relatively stronger than those between other GRE item-type scores. Each GRE subscore tended to have higher correlations with the GRE item-types constructing the subscore than with GRE item-types constructing other test subscores. The analysis of correlations among GRE item-types shows that the item-types have strong internal validity.

Wilson (1985) has suggested that GRE (and SAT) item-types may measure discrete forms of general education abilities. This assertion served as the theoretical underpinning for the use and treatment of GRE item-types as discrete, multiple measures of general learning. To test Wilson's assertion, the intercorrelation among item-type scores was further examined (see Figure 4-3).

In the Evergreen State Sample, intercorrelations for Verbal item-types ranged from r = .35 (RD/ANA) to r = .67 (ANT/RD). Intercorrelations for Quantitative item-types ranged from r = .48 (QC/DI) to r = .64 (RM/QC). Intercorrelations between Analytic item-types were r = .59 (ARE/LR). However, Analytic Reasoning correlated strongly with Quantitative item-types ranging from .39 (DI) to .67 (QC). The intercorrelational analyses showed that in all instances, less than 50 percent of the variance in one item-type was explained by that of another.

Figure 4-3. Intercorrelation of GRE Item-Types--Evergreen State Combined Sample

GRE Item-Types	Code	ANA		SC		RD		ANT		QC	
Analogies	ANA	1.0000									
Sentence Completion	SC	.4333	[2]	1.0000							
Reading Comprehension	RD	.3477	[1]	.4422	[2]	1.0000					
Antonyms	ANT	.3908	[2]	.4625	[3]	.6645	[4]	1.0000			
Quantitative Comparison	QC	.4131	[2]	.2721		.5321	[4]	.4634	2	1.0000	
Regular Mathematics	RM	.3442	1	.1816		.3868	[2]	.1526		.6377	[4]
Data Interpretation	DI	.2594		.0614		.1115		.1162		.4764	[3]
Analytic Reasoning	ARE	.4942	[3]	.4762	[3]	.6574	[4]	.5120	[3]	.6735	[4]
Logical Reasoning	LR	.5685	[4]	.4394	[2]	.5497	[4]	.4336	[2]	.5377	[4]

GRE Item-Types	Code	RM		DI		ARE	LR
Analogies	ANA						
Sentence Completion	SC						
Reading Comprehension	RD						
Antonyms	ANT						
Quantitative Comparison	QC						
Regular Mathematics	RM	1.0000					
Data Interpretation	DI	.6013	[4]	1.0000			
Analytic Reasoning	ARE	.5949	[4]	.3870	[2]	1.0000	
Logical Reasoning	LR	.4844	[3]	.3028	[1]	.5911 [4]	1.0000

[1] $p < .05$ [3] $p < .001$

[2] $p < .01$ [4] $p < .0001$

As Figure 4-4 demonstrates, Evergreen State students performed well on the GRE General Examination. Students answered the questions correctly in approximately 113 of the 186 GRE items. A few students attained perfect scores on Sentence Completion item-types.

While GRE raw scores were generally and consistently high among these students, differences among scores appeared when the effect of the precollege learning (as measured by the SAT) was removed. When the theoretical scores (as predicted by corresponding SAT scores) were compared with the students' actual

responses (Figure 4-5), students showed the largest improvement on the Data Interpretation item-type and the lowest amount of improved performance on the Quantitative Comparisons item-type.

The greatest amount of variance in item-type residuals, including the greatest standard error and standard deviation, were found in the Analytic Reasoning and Quantitative Comparisons item-types. The variance in these residuals holds implication for the ensuing cluster analysis in that GRE item-types with greater variance will play a more significant role in sorting courses into clusters. As was discovered in the analysis of samples from other participating institutions, those GRE item-types with smaller variance play less of a role in discriminating course clusters.

As Figure 4-5 demonstrates, from one-eighth (Data Interpretation) to two-thirds (Quantitative Comparison) of GRE item-type score variation among the Evergreen State Sample was explained by their SAT scores. All regression functions were statistically significant at .0001 with the exception of Data Interpretation which was significant at .006. Also, the range of residual scores did vary considerably across GRE item-types.

Using the student residuals obtained from the regression analysis above, the mean residuals for each program enrolling 2 or more students were calculated for all the 9 GRE item-types. Such a procedure does not assume that the specific improvement of the students enrolled in each program were directly caused by that course. Rather, the residuals of each student are attributed to all the programs in which they enrolled, and the mean residuals for each program serve as a proxy measure of student gains. Once programs are clustered by these gains, then hypotheses can be generated and tested as to why students who enrolled in a given pattern of programs experienced significant gains on one or more of the outcomes criteria (i.e., the item-type residuals).

Figure 4-4. The Distribution of GRE Scores for Evergreen State Combined Sample

GRE Item-types	Number of Items	Minimum Right	Maximum Right	Score Range	Sample Mean	Standard Deviation
Analogy	18	7	16	9	11.82	2.3693
Sentence Completion	14	5	14	8	10.76	2.3398
Reading Comprehension	22	7	20	13	13.71	3.2841
Antonyms	22	6	21	15	13.58	3.3769
Quantitative Comparison	30	10	27	17	18.21	4.8301
Regular Mathematics	20	4	19	15	11.59	3.3703
Data Interpretation	10	0	8	8	5.00	1.6997
Analytical Reasoning	38	16	35	19	23.67	5.4734
Logical Reasoning	12	3	11	8	7.26	2.2453
GRE Verbal	76	32	67	35	49.89	8.8650
GRE Quantitative	60	18	53	35	34.80	8.5469
GRE Analytic	50	20	46	26	30.93	7.0377
GRE Verbal (converted)					548.00	91.1442
GRE Quantitative (converted)					506.67	106.4297
GRE Analytic (converted)					572.00	100.3766
Minimum	10	0	8	8	5	1.70
Maximum	38	16	35	19	23.67	5.47
Mean	21	7	19	13	12.97	3.33
Total	186				115.60	28.99

Figure 4-5. Summary of Regression Analysis of GRE Scores--Evergreen State
==

Dependent Variables	Evergreen State Sample 46 Students		
GRE Item-types on SAT Sub-scores	Code	F Value	Standard Deviation	Adjusted R-Squared
Analogies	ANA	36.774	2.3692	.4429
Sentence Completion	SC	35.973	2.3398	.4373
Reading Comprehension	RD	22.385	3.2841	.3221
Antonyms	ANT	48.185	3.3768	.5119
Quantitative Comparisons	QC	83.045	4.8301	.6458
Regular Mathematics	RM	25.914	3.3703	.3564
Data Interpretation	DI	8.2710	1.6997	.1391
Analytic Reasoning	ARE	44.374	5.4734	.4908
Logical Reasoning	LR	34.754	2.2453	.4286
Verbal (raw)	GRE-V	112.52	8.8650	.7125
Quantitative (raw)	GRE-Q	66.636	8.5470	.5933
Analytical (raw)	GRE-A	61.322	7.0377	.5727

$p < .0001$

==

Regression analysis of SAT scores on GRE item-type scores

To determine the extent to which these students showed improvement over
their precollege SAT scores, the GRE raw scores were regressed on the
corresponding SAT scores. GRE Verbal item-types were regressed on SAT Verbal
sub-scores, GRE Quantitative item-types were regressed on SAT Math scores, and
GRE Analytic item-types were regressed on SAT Total scores. The resulting GRE
item-type correlations with corresponding SAT scores were noticeably lower than
their corresponding GRE sub-scores. This suggests that individual item-types
also may measure discrete abilities apart from those of the SAT sub-scores
and/or they may reflect lower reliability stemming from the fact that there are
fewer items comprising an item-type than a sub-score.

The SAT scores explained smaller portions of variance in GRE item-type scores than in the GRE sub-scores (Verbal, Quantitative and Analytical--see Figure 4-5). The SAT Verbal explained 43.73 percent of the variance in the Sentence Completion item-type among the Evergreen State Sample. The SAT Verbal explained 44.29 percent of the variation in the Analogies item-type, 32.21 percent of the variation in Reading Comprehension and 51.19 percent in Antonyms items among the Evergreen State Sample. The SAT Math scores explained 64.58 percent of variation in Quantitative Comparison item responses, 35.64 percent of variation in Regular Math item-type scores, and 13.91 percent of variation in Data Interpretation for the Evergreen State Sample. The combined SAT Verbal and SAT Math scores (referred to as SAT Total) explained 49.08 percent of variance in Analytic Reasoning and 42.86 percent of variance in Logical Reasoning for the Evergreen State Sample. In all but one instance, the regression model proved significant at the .0001 level, suggesting effective control measures for the general learned abilities of students as they entered college as freshmen. The only exception was Data Interpretation which was significant at the .006 level.

Evergreen State Sample students entered college with slightly higher mean SAT Verbal score (529) than SAT Math score (513). As these Evergreen State students approached graduation, they remained normatively stronger in Verbal abilities (mean GRE-V = 548) than in Quantitative abilities (mean GRE-Q = 507). Yet, the regression analysis showed the Evergreen State students evinced large variance in residuals on specific GRE item-types in excess of those represented in the sub-scores of the two tests. For Evergreen State, specific and significant residual variance was demonstrated in Quantitative Comparison and Analytic Reasoning.

Figure 4-5 compares the explained variance (R-squared) for each GRE item-type and raw GRE sub-score. Only SAT sub-scores were available to the

research team; these scores are <u>converted scores</u>. The actual scores of a student on a particular form of the SAT test is transformed relative to test norms so as to be comparable with other forms of the test and with other students tested. A similar process is used with the GRE exams. Raw GRE sub-scores for a given form of the test are transformed so as to allow comparisons with national norms and with scores on other forms of the test. In all cases within the Evergreen State Sample, the SAT accounted for more variance in GRE sub-scores than in the GRE item-type scores. Figure 4-6 illustrates the extent of <u>unexplained variance</u> (that is, the varia..ce in GRE item-type scores attributable to sources other than the precollege abilities of the students, as measured by the SAT). Only the raw GRE sub-scores are graphed in Figure 4-6. The findings tend to agree with previous reports of this project and those by Wilson (1985) suggesting that GRE item-types may have greater correlation with variance in learning during the college years than do the GRE sub-scores.

As has been previously discussed, critics of the GRE and SAT as measures of general learned abilities attack the validity of the measure themselves. These criticisms are based primarily on the use of the te't sub-scores and the total test scores; the use of the item-type scores on either the GRE or SAT as <u>multiple measures</u> of general learning have not been widely explored (Adelman, 1988). The reliability of GRE item-types, their strong correlation with SAT sub-scores, and their apparent ability to measure discrete types of learning suggest that they may be of potential value as criteria in the assessment of general learning abilities of undergraduates.

For the purposes of this research, precollege measures of student ability were defined as the SAT-V, SAT-M, and the total SAT score. Although there is research to suggest that SAT item-types may have greater predictive validity of college performance (Ramist, 1981a; 1981b; Schrader, 1984), SAT item-types were

not available for use in the research model. Postcollege measures of student learning were defined as the 9 item-type scores on the GRE. According to Astin's model (1970a, 1970b), the effect of the learned abilities of students entering college on student outcome measures should first be determined. Once the variance in GRE item-type scores attributable to SAT scores was determined, the unexplained variance (score residuals) could be used as a proxy of change in general student learning along the 9 item-type measures.

The SAT Verbal scores were used to predict each of the GRE Verbal item-types using the general lineal model. The SAT Math scores were used to predict each of the corresponding GRE Quantitative item-types. The combined SAT scores (SAT Total) were used to predict the GRE Analytic Scores. The regression analysis summarized in Figure 4-5 was performed by the PROC REG in the SAS statistical package. Since individual scores were predicted relative to the sample group, some students will have higher actual scores than predicted scores, while others will have lower actual scores than predicted scores. A negative residual represents the unexplained variance of a student whose actual GRE item-type score was less than that predicted by the student's corresponding SAT subscore. Larger unexplained variance on a given item-type indicated larger amounts of change (either gain or loss) in general learned abilities associated with that measure.

Variation in GRE item-type scores may be attributed to two sources: variation due to changes in the independent SAT variable and changes attributable to other sources. In most cases in the Evergreen State Sample, the probability of obtaining the F value was .0001, suggesting that the general linear model is adequate in explaining the sources of variation within the GRE item-type scores. R-squared represented the percent of total GRE item-type variation explained by the independent SAT variable. The residual variation in

- 82 -

each GRE item-type--that not measured by the SAT score--was used as the proxy measure of general learned abilities during the undergraduate experience. These student residual scores, from the time they took the SAT prior to college to the time they took the GRE as graduating seniors on each of the 9 item-types, served as the measures of general learning among the Evergreen State Sample (Hanson, 1988; Pascarella, 1987).

Fig 4-6. Change by GRE item-types

Evergreen State

and GRE sub-scores
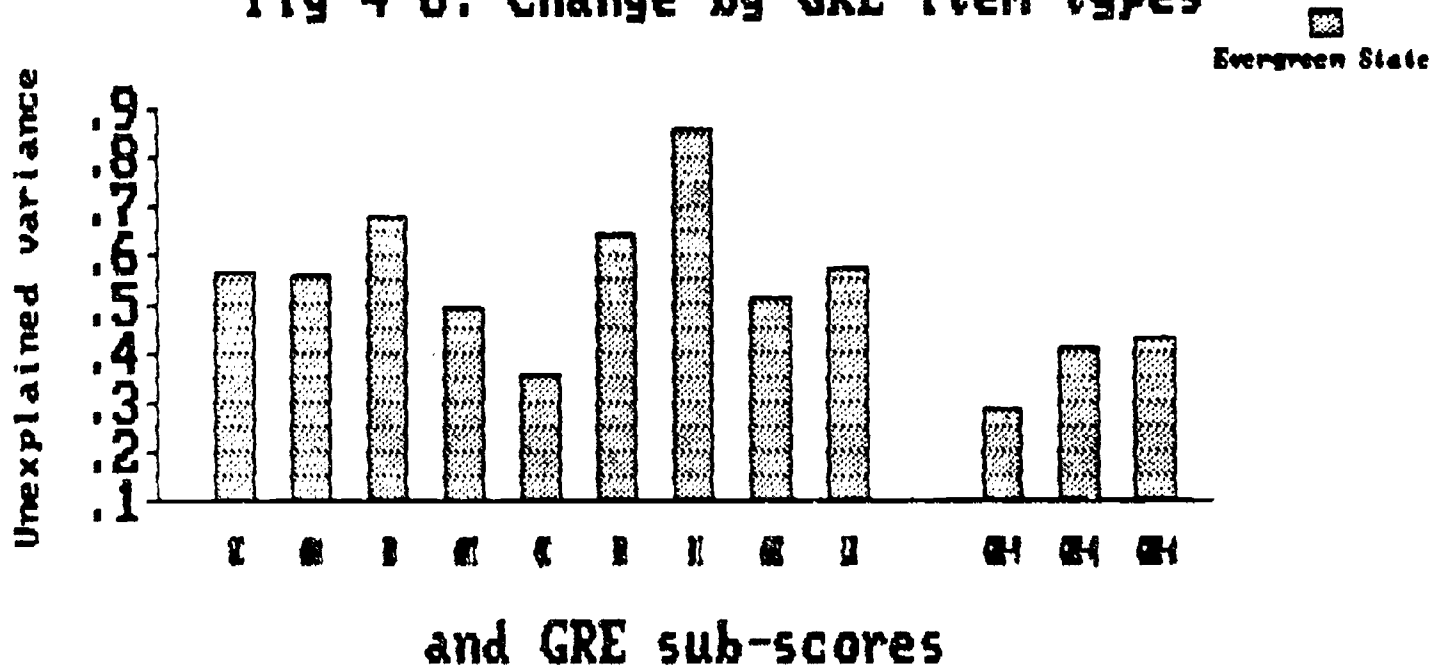
V. <u>The Evergreen State Catalog Study</u>

<u>The Use and Limitations of Catalog Research</u>

Transcripts provide a variety of information which is keyed to a coding system or scheme used by the college or university. Transcripts often provide short or abbreviated titles of courses, degrees, majors. The coding scheme colleges and universities use often is similar at the macro level and Byzantine in its complexity at the micro level. Certain conventions appear to guide in the abbreviation of departments and programs. English, for example, is usually designated at ENG or ENGL on transcripts, except where further differentiation is sought (LIT for literature or COMP for composition courses). An ultimate aim of the transcript is to present the student's program of study to the outside world in some intelligent fashion (Andrews, 1985; Dressel & DeLisle, 1969).

Evergreen State College transcripts are quite different from these conversions. For each course or program in which a student enrolls, two or three narratives are generated. First, at the outset the supervising faculty writes a description of the purpose, content and procedures of the program or course. Second, at the conclusion of the program, the faculty describes the student's achievement and accomplishments. Thirdly, the student may also write into the transcript record their own perceptions of what they learned as a result of the program or course. Such narrative provides a rich detailed reflection of eacn student's learning. Forty-page transcripts of such learning are not uncommon for Evergreen State.

Looking at a single, conventional transcript or paging through a college catalog gives a false sense of simplicity to course numbering and labeling. There is usually some designation as to which department or division within the institution offered the course, but it is not unusual to find departments with

courses designated in two or more ways. For example, on a transcript may appear ANTH 190, "Dances of Other Cultures," offered for 2 credits. The designation ANTH 190, by convention at many colleges, indicates a freshman level course in anthropology. However, through the examination of several other courses, we may find ANTH 390, "Dances of Other Cultures," indicating a junior level course in anthropology for 3 credits. Indeed, further examination of transcripts also reveals a DANCE 190 which has the same course title for 2 credits and DANCE 390 with the identical title for 3 credits. Referring to the college catalog, we discover that ANTH 190 and DANCE 190 are the same course, cross-listed by both departments to indicate that students majoring in either subject may count the course toward completion of degree requirements within the major. Furthermore, we also discover that ANTH 190 and ANTH 390 are basically the same course (as are DANCE 190 and DANCE 390), but provision has been made in DANCE 390 and in ANTH 390 for extended study and research for students enrolling in the course as upper division students. Additional examination of transcripts may also reveal that ANTH 190 was first offered in 1980, that it was cross-listed as DANCE 190 in 1982, and that ANTH 390 and DANCE 390 were added to the curriculum in 1984. Looking at transcripts of graduating seniors in 1987, we may encounter all four course designations. To determine the effect of coursework patterns on general learned abilities, we must first decide what course designations are in fact equivalent and which courses are discrete learning experiences. To make that determination, the college catalog is a limited as a useful tool.

The purposes of the Evergreen State College Catalog Study were two-fold. First, we wanted to determine cross-listed programs and programs that were otherwise equivalent to each other. Second, we needed to learn more about individual programs than their short program titles indicated. To know that a program is offered by the English faculty tells us little as to whether the

intent of the program is to teach business composition (skill development), literary criticism (methodology), or a survey of literature (content, history). We first examined the purpose, procedures and findings of prior research on college catalogs.

Review of related research

The review of selected literature on research in which college and university catalogs served as sources of data and analysis revealed three basic areas of investigation of the higher education curriculum. One area of curricular research used catalogs to analyze the development of a specific academic area or a specific topic. The second area of research which employed catalogs studied the change in the general higher education curriculum that occurred over a specific time period. A third research area identified in the literature involved the use of catalogs to develop course classification systems. In all three types of investigation, course descriptions, course titles, and degree requirements provided most of the data.

Catalog studies of specific academic areas

The specific academic disciplines studied included acting, medical technology, information science, technical communications in engineering, and higher education administration. The specific topics studied included the presence of a global/international perspective in the curriculum of several colleges and universities, and the use of the catalog as an aid in career decision-making.

Becan-McBride (1980) studied the catalogs and brochures from 367 medical technology programs to determine: (a) the prerequisite courses needed for admission to the medical technology programs, (b) the course requirements for these programs, (c) the type of program, (d) the nature of the program

- 86 -

1(·1

accreditation, (e) fees and credits required, (f) the number of credits earned in the senior year, and (g) the type of degree earned. Programs were classified into hospital-based and university-based curricula and were cross-tabulated with the prerequisite listings. The frequencies of various program prerequisites and admissions prerequisites were cross-tabulated by type of program. A chi-squared was computed to determine if the two types of programs were significantly different in program prerequisites. The results indicated that four courses appeared more often in the university-based prerequisite courses than in the hospital-based programs list of prerequisite. Becan-McBride interpreted this difference to indicate a lack of consensus between university-based and hospital-based medical technology programs as to the desired prerequisites for admission. McBride's study suggests prerequisites to be significant variables in the examination of catalogs. Furthermore, the understanding of prerequisites in the college catalog may help identify constraints upon student course-taking behavior as evidenced on transcripts.

Tenopir (1985) reviewed college catalogs in which information science or related words appeared in the course descriptions, course titles, and/or department titles. The purpose of the study to discover what kinds of college courses determined a specialization in information science. The researcher chose to use catalogs because they constituted unobtrusive measures of what faculty deemed to be an area of specialization. The study was limited to the most recent catalogs for each college studied. From the catalog study, Tenopir used the 22 categories of Information Science courses developed by Blezer (1971) using cluster analysis and Delphi techniques. Each course was assigned to only one category although many courses overlapped several categories. The placement of courses in a category was based on the catalog course descriptions.

Tenopir found that masters-level courses dominated the information science course offerings. "Library Science" was the dominant parent department for offering information science courses. The researcher noted, "that the word `information' did not appear in over 60% of the department titles" which "clearly points out the problems of identifying schools teaching information science by looking for the word 'information' in their titles." Course lists were made for each parent department. The departmental course lists demonstrated that many of the same courses existed in all the parent departments. Tenopir acknowledged that her study was delimited to the catalogs stud ed and reflected only a small portion of the information science courses offered in the United States. However, she cautioned that an examination of all American college and university catalogs to identify and classify course offerings would not be justified by the time and labor involved. As an alternative, she suggested that future researchers examine only those college departments who appeared in information science directories or only those college and university catalogs that use "information science" in the title of a department. Tenopir noted that using these limitations would exclude some college departments from study but it would produce a profile of departments that claim to teach information science. Tenopir's study illustrated many of the problems attendant to the lack of a common nomenclature for college catalogs.

Boysen (1979) analyzed the technical communications curriculum for undergraduates at selected colleges of engineering. The researcher used college catalogs for the study. Two of the criteria used to select the colleges that participated in the study were derived from the college's catalog. Only those programs with clearly written course descriptions and lucidly composed program descriptions were included in Boysen's study. Boysen found that most colleges recommended that engineering students enroll in the technical communication

classes in their junior or senior year. Boysen's study showed that many
catalogs are replete with ambiguities; it also suggested that the intent of the
curriculum can be comparable where catalogs are clearly composed and the
programs of study are similar.

LoGuidice (1983) analyzed the contents of 34 catalogs from Lutheran-
affiliated colleges and universities to determine the presence of a global/
international perspective in their curricula. He analyzed the content of course
titles and courses descriptions, looking for the a prior selected words
"global," "international," and "comparative studies". Courses considered to be
regional studies, language courses, and general liberal arts courses were not
included in the study. The study found that the most commonly used terms in the
catalog listings and descriptions were "international affairs" and
"international trade." The researcher also found that the courses using the
phrase "global perspective(s)" or "international perspective(s)" occurred most
in economics and political science departments. LoGuidice, like other catalog
researchers, used simple frequency counts of predetermined words or phrases to
examine the presence and prevalence of specific characteristics or content of
the curriculum. As in other studies reviewed, LoGuidice found commonalities of
wording among like departments or disciplines.

Studies of catalog change over time

Gillespie and Cameron used college catalogs, textbooks, and national
convention programs to determine the development of the teaching of acting in
American higher education from 1920 to 1960. Hamar (1951) also used catalogs to
study acting education. The catalogs of 58 colleges were used in the Gillespie
and Cameron study; they also developed a more precise definition of "acting" for
their use in analyzing the catalogs than did Hamar. They did this to separate

- 89 -

104

acting classes from closely related subjects such as oral interpretation and preparation for plays. Therefore, they chose only courses in which acting or a variation of the word acting appeared in the course title or course description.

Gillespie and Cameron (1986) reported the appearance of acting classes in catalogs increased from 1920 to 1960. The researchers felt this meant that acting gained acceptance as a college course over the forty years they studied the catalogs. Additionally, they found specialization in acting education increased from 1920 to 1960. A dispute over the teaching methods of acting appeared in catalogs studied by the researchers. Some colleges taught acting by participation in productions and other colleges taught acting mostly in classroom situations. Another result showed that the Stanislaski approach to the teaching of acting appeared much later in the catalogs' course descriptions than it did in the acting textbooks and in the acting profession.

Fosdick (1984) made two surveys of graduate library schools catalogs. One occurred in 1977 and the other in 1982. The researcher used the survey results to determine the effect of information science courses in the library science curriculum. The 1977 catalog survey included 54 library schools, while 62 schools were investigated in the 1982 catalog study. The purpose of these analyses was to determine the educational impact of information science on the library science curriculum. Course titles and descriptions were used as sources of information because Fosdick believed them to be more objective than verbal or written questionnaires. According to the researcher, "... the schools cannot tailor their responses to the aim of the survey." However, Fosdick found that the analysis of catalogs also presented some problems:

> investigation of the catalogs does involve the subjective judgements of the surveyor, and a few schools do not provide adequate information in their catalogs on which to base an appraise ... (p. 293).

Fosdick's 1977 and 1982 catalog studies found that information science courses fell into five main groups. The number of information science courses offered in the library science graduate schools increased from 1977 to 1982. This meant, according to the researcher, that the study of information science increased in import ance in the library science graduate schools.

Grace's 1985 study examined the content of the higher education administration curriculum in order to determine the extent of professionalization within the field of higher education administration. She reviewed other professions to develop a basis for deciding the degree to which higher education administration had matured as a profession. The higher education administration programs of Indiana University, Teachers College-Columbia University, and the University of Michigan participated in the study. For her study Grace used the 1972-73 catalogs, 1982-83 catalogs, student program handbooks, and student trans ripts from each program. These data were categorized into grids used in Blackburn's 1976 study. Grace used the college catalogs to gather the following data items: total program hours required, number of recommended courses in higher education, number of recommended courses not in higher education, the areas of recommended courses not in higher education, and requirements according to area. To analyze the characteristics of the higher education curricula, Grace used a model developed by Conrad (1978). This model applied a priori continua in four areas to examine curriculum: (a) locus of learning, (b) depth/breadth, (c) design of program, and (d) flexibility of program. Using Conrad's model, Grace also analyzed four other professions: law, medicine, library science, and business. These were compared to higher education administration. The researcher indicated that higher education was similar to the other professions in its stages of professionalization. She recommended that future research include the analysis of course-taking behavior in the four professions compared in order to

- 91 -

106

identify relationships in course-taking behavior among higher education administration and the other four professions.

Dressel and DeLisle (1969) studied general higher education curricular trends and changes over a 10- year period from 1957 to 1967. They chose this time/frame because of its frequent reference and common identification in the literature and at conferences as a time of great curricular change. Dressel and DeLisle examined the 1957 and 1967 catalogs of 322 colleges and universities selected at random from the American Council on Education's 1964 edition of American Universities and Colleges. The researchers determined change on the basis of the depth (percentage of courses needed for fulfilling requirements for a major) and breadth (percentage of courses needed for fulfilling general education requirements) of the curriculum. Dressel and DeLisle found particular kinds of curricular change were not distinct to specific institutional type. They reported that the general education requirements of most colleges and universities became less specific in 1967 than had been the case in 1957. They found course and credit requirements for an academic major varied widely. However, it was difficult to determine the extent of major requirements because of the many differences among the colleges and universities in defining an academic major. They found the use of catalogs as a source of information for a study created some problems because the catalogs were poorly organized, and sometimes contradictory and inaccurate.

During the period of the study, Dressel and DeLisle also noted an increase from 1957 to 1967 in the proportion of elective courses a student could apply to the fulfillment of their degree requirements. They suggested caution, however, in drawing conclusions regarding the use of catalogs to determine the proportion of elective courses in a program:

> [the] exact proportion of a program available for individual choice
> within distribution requirements is not entirely free, and it may
> further delimited by advice or by major requirements. Choice of a
> major and of courses within a major may be much less than free, for
> the student usually must have a major, and his courses within the
> major ... may then be directed, if not dictated (p. 43).

Dressel and DeLisle also reported an increase in the individualization of the students' learning experiences from 1957 to 1967. Individualization would include such courses as advanced placement, honors programs, independent study, seminars, and study abroad.

Hefferlin (1969) examined curricular change in American higher education over a 5-year period from 1962 to 1967. The focus of his study was on the expansion and reform of the curriculum. The researcher wanted to determine what new courses had been added to curriculum and how the content of the courses changed over the 5-year period of the study. Hefferlin studied the bulletins and catalogs of 121 colleges and universities in a stratified random sample of the 1966-1967 Education Directory of the United States Office of Education. The sample was stratified according to the number of institutions in each of the 48 contiguous states.

Hefferlin used catalogs to study five areas of curricular change: (a) courses offered, (b) program majors and areas of concentration, (c) degree requirements, (d) requirements within a major, and (e) general regulations regarding the curriculum. Hefferlin and his associates studied four departments at each institution. The departments were chosen randomly -- one each from humanities, behavioral sciences, natural science, and vocational subjects. Hefferlin (1969) measured the percentage of reform by determining the proportion of courses that changed noticeably or were dropped during the 5 year period in departments where the number courses increased. He also inspected the proportion of courses that had been added or changed in departments where number

courses in the department declined.

Based on these measures of change within the curriculum, Hefferlin created
a model for studying curriculum reform, The Study of Institutional Vitality
(SIV). Hefferlin's model isolated features that differentiated between
institutions that were static and those that were dynamic. This measure of
vitality correlated information important to the study of colleges and
universities to the facts about academic change in colleges and universities.
The hypothesis that the researcher developed, from the existing evidence on
academic change, was that change occurred because of organizational
instability. The instability was caused by problems with resources, staff, and
structure. Hefferlin said, "the most evidence of curricular change would be
found at the weakest, most marginal, most down and out colleges, where perennial
crisis stimulates the most frantic adaptability." When the factors were
correlated, the researcher found that academic reform was a function of
organizational instability.

Blackburn, Armstrong, Conrad, Didham and McKune (1976) conducted a study of
undergraduate education for the Carnegie Commission on Higher Education. The
purpose of the investigation was to answer two questions regarding the status of
undergraduate education. First, the study asked how much change occurred in the
breadth and in the depth of the curriculum from 1967 to 1974. The breadth was
defined as a percentage of the undergraduate degree requirements which were
represented by general education courses. The depth of the curriculum was
defined as the percentage of courses that met the requirements for an academic
major. A second question asked how the course taking behavior of the students
related to the degree requirements from 1967 to 1974. The study consisted of
two phases based on these two questions.

In the first phase of their study, Blackburn et al. examined the courses designed to fulfill general education and major requirements. The researchers used the 1966-1967 and 1973-1974 catalogs from 271 colleges and universities. The sample colleges and universities were selected on the basis of the Carnegie Foundation's classification of different kinds of colleges and universities (Carnegie Council, 1976). The analysis of the catalogs indicated that the proportion of general education courses required for an undergraduate degree decreased from 1967 to 1974 and that the proportion of courses required to fulfill requirements of the undergraduate major changed very little from 1967 to 1974. These trends resulted in a net increase in the elective courses students were able to include in their baccalaureate program of study.

The second phase of the Blackburn study examined the course-taking behavior (as evidenced by transcripts) of students from 1967 to 1974 and compared the course-taking behavior with the degree requirements (as evidenced by the catalogs) during the six-year time period. The researchers selected ten institutions for participation in this second phase of the study. Eight of the institutions experienced some curricular change from 1967 to 1974. The researchers defined curricular change as when the institution reduced its general education requirements and/or shifted general education requirements from a list of specific courses to a less specific group of courses from which the students could choose to meet their general education requirements. Two of the colleges and universities were selected as a control group because they had not experienced any significant curricular change. The researchers analyzed the transcripts from a random sample of 30 different Bachelor of Arts degree recipients (representing six academic majors) for each year of the study. Transfer students from other colleges and universities were excluded from the study. The researchers found that students' course-taking behavior generally

agreed with degree requirements stated in the catalogs. However, it has been noted elsewhere (see Chapter 1) that Blackburn's investigation was based on an extremely small sample of courses within the curriculum by virtue of the small number of transcripts examined.

Blackburn et al. (1976) reported this study to be the first time where transcripts were used for the analysis of curriculum change. Transcripts allowed for monitoring of practices that could not be determined from catalogs. However, the researchers suggested the use of transcripts be limited because it was a slow and expensive process. Blackburn et al. called for additional research to determine if the different teaching techniques and curricular programs lead to any difference in the learning outcomes.

Studies of course coding and classification

In a 1985 article Andrews explained how Illinois Valley Community College (IVCC) revised its catalog and reported the results of a survey of high school guidance counselors' evaluations of the revised catalog. The IVCC catalog revision began with a college decision to create a more innovative catalog than what had produced for many years. An analysis of the current catalog suggested that it lacked clarity in what the college could offer potential students. Andrews found that the old catalog presented information that students did not need to know until they enrolled in the College. The new catalog would emphasize the vocational interests of students. Each curricular program was renumbered to correspond with two career-interest instruments frequently used by local high school counselors and the college's counselors. The rationale for the change in program coding was based on the fact that more than 33% of the incoming IVCC students indicated they were unsure of their career goals. The new curriculum coding system was based on Holland's Theory of Vocational

111

Development. This theory corresponds with two career-interest instruments used
by the counselors, the Strong-Campbell Vocational Interest Inventory and the
Holland Self-Directed Search. The survey of high school counselors indicated
the counselors thought the revised catalog's coding system made the new catalog
much more useful than earlier catalogs. The IVCC catalog revision suggests that
a gap may exist between the departmental coding traditionally used in catalogs
and the career interests of college-bound students.

A Classification of Secondary School Courses (1982) and Waggaman (1980)
used catalogs to study course classification systems. The former focused on a
nationwide system and the latter on a statewide system.

The Classification of Secondary School Courses (1982) project used course
descriptions collected from high school course catalogs to develop a nationwide
inventory of secondary school courses. The directory, Classification of Second-
ary School Courses (CSSC), would be used by the National Center for Educational
Statistics (NCES) for coding high school transcripts. The project collected
10,000 course titles from a sample of 525 catalogs of the 1979-1980 and 1980-
1981 academic years. The schools were selected from a sample of schools used in
National Longitudinal High School and Beyond Study (NLS). The sample included
private religious and independent schools, and public schools. The classifi-
cation was arranged according to the Classification of Instructional Program
(CIP). This classification system was developed by NCES for post-secondary
education. A panel of reviewers examined the catalogs and established a course
title index for each course (using the CIP system), a unique 6-digit code
number, keyword descriptors, and alternate course titles. The reliability of
the coding system was tested by four coders who were given directions about how
to make coding decisions and 50 course catalogs to code courses. Waggaman
(1980) examined the effect of varying forms of credit and non-credit

- 97 -

112

designations on the Florida Statewide Course Number System (SCNS). The expansion of the kinds and uses of credit courses created the problems for the State-wide Course Number System (SCNS) committees. The SCNS committees classify new credit offerings to establish a common course numbering system for all courses offered in Florida's public colleges and universities. The project studied college and university course catalogs (1979-1980 and 1980-1981) from Florida's institutions of higher education, various state and federal laws, and accreditation documents from professional organizations. The analysis of this information led to an examination of the past, present, and future practices regarding the use of credit in Florida and recommendations for specific policies to reduce the present problems in classification of courses by level and department.

Discussion and results of the literature review

A common research design was used in the above studies. In general these studies either derived from prior literature or established de novo predetermined categories of courses. Only Blezer (1971) provided an empirical basis for course categories. They conducted simple word counts or qualitative content analyses of the course titles and/or course descriptions to determine the frequency of courses occurring in each category. These studies did not evidence any formal procedures to examine the content or concurrent validity of the categories used. Several studies, (Boysen 1979, Fosdick 1984 and Tenopir 1985 notably), reported difficulty in placing courses into the predetermined categories and identifying like programs (Dressel & DeLisle, 1969). Ambiguities in wording and phraseology troubled the catalog analyses. Several researchers-responded by erecting more discrete and exacting operational definitions upon which to judge and classify courses (Boysen, 1979; Gillespie & Cameron, 1986).

Again, the validity and reliability of the classifications limited to clearly-worded courses titles and descriptions has not been examined.

Similarly, these studies made little effort to determine inter-rater reliability in reading and categorizing the catalog course titles and descriptions. While Becan-McBride (1980) used Chi-square to examine the distribution of these counts among categories and between types of institutions, most researchers were content to amass and generalize from simple frequency counts.

A major attraction of researchers to catalogs as sources of information was the absence of subjectivity and intrusiveness attendant to survey questionnaires. Yet, the analysis and classification of courses, programs, requirements and degrees found in catalogs was found to depend upon the subjectivity of the research performing the study (Fosdick, 1984).

The review of the above catalog studies of specific academic areas identified several methodological and substantive issues in using catalogs as unobtrusive measures of college curriculum. First, prerequisites appear as variables posing potentially significant constraints on course-taking behavior (Becan-McBride, 1980). Second, course content may exhibit commonalities across institutions by academic department or program (Dressel and DeLisle, 1969) and among institutions by type (Becan-McBride, 1980; Boysen, 1979; LoGuidice, 1983; Tenopir, 1985). However, general education requirements may be subject to broader social and intellectual trends and may vary regardless of institutional type (Dressel & DeLisle, 1969). Third, the number of hours required of a specific program, the number of recommended and the number of required courses within a field of study and outside a field of study were also found to be potential constraints on study course selection and enrollment (Grace, 1985). Fourth, the proportion of elective courses and individualized study were found

to vary across institutions over time (Dressel & DeLisle, 1969).

Studies of catalog changes over time suggested that courses were added to the curriculum as new subjects or techniques were added to a given field of study (Gillespie & Cameron, 1986; Fosdick, 1985). In the five year-span analyzed by Fosdick, there was significant growth in numbers of courses and subjects within the field of study-library science. An analysis of transcripts over this same time span would presumably be influenced by the proliferation of courses in the curriculum. Dressel and DeLisle (1969) noted changes in general education, electives and individualized instruction opportunities over a ten-year period of rapid curricular change. Grace (1985) provided an example of how course-taking behavior could be contrasted to stated program requirements and guidelines according to an externally-derived set of criteria. Blackburn et al. (1976) provided evidence that the number of general education requirements and requirements within the major had declined over the 5-year period of 1967 to 1974, resulting in a net increase in elective courses that could be potentially included in a student's program of study. These findings, that the number of courses increased and that elective and individual instruction opportunities have increased during the late 1960s and early 1970s, demonstrated how broad curricular changes could liberate or constrain students' course-taking behavior.

Hefferlin (1969) measured the percentage of reform by determining the proportion of courses that changed noticeably or were dropped during the 5-year period in departments where the number courses increased. He also inspected the proportion of courses that had been added or changed in departments where number of courses in the department declined. Blackburn, Armstrong, Conrad, Didham and McKune (1976) asked how much change occurred in the breadth and in the depth of the curriculum from 1967 to 1974. The breadth was defined as a percentage of the degree requirements for an undergraduate degree which were represented by

general education courses. The depth of the curriculum was defined as the percentage of courses that met the requirements for an academic major. These two studies provided models for interpreting the simple frequencies of courses in the catalog over time into measures of institutional vitality and stability and into measures in the basic structure of degree requirements.

The curriculum at Evergreen State College

Evergreen State College is a liberal arts and sciences college with some very special features. Located just outside Olympia, Evergreen was created to serve as a regional learning center for the citizens of southwest Washington and as an educational alternative to the state's other colleges and universities. Evergreen also serves as an educational and research resource for state government. Since it opened in 1971, Evergreen has distinguished itself by the accomplishments of its faculty, students, and graduates, and through its special approach to education.

Unified and interdisciplinary programs

The curriculum is organized into interdisciplinary, unified and focused programs of study. All of the student's undergraduate program is coordinated around a central theme or issue. Although all Evergreen programs tend to embrace a number of different teaaching methods, seminars are the dominant mode of instruction. Reading, writing, discussion, and research all develop the program theme. Learning to make those connections is one of the larger purposes of education at Evergreen. This interdisciplinary approach to studying one topic from different disciplinary perspectives is intended to allow students to integrate diverse pieces of knowledge into a larger framework.

Credits and Evaluation

Most programs represent a full academic load of 16 quarter hour credits. The minimum requirements for awarding either the Bachelor of Arts (B.A.) or the Bachelor of Science (B.S.) is 180 quarter credit hours. The B.S. degree requirement also includes 72 quarter credit hours in mathematics and natural science, of which 48 quarter credit hours must be in advanced subjects. Concurrent award of a B.A. and B.S. requires a minimum of 225 quarter hours, including a minimum of 90 credit hours earned in residence at Evergreen and application at least one year prior to admission.

Evergreen recognizes that students differ in maturity and personality as well as interests and capabilities. The college acknowledges and tries to build upon their diversity. Therefore, there are no structured majors or specific required courses. The alternative to requirements is not random choice of academics, but rather a highly individualized, carefully thought-out educational plan.

Instead of consisting of a mere list of course titles and letter grades, a student's transcript is a compendium of faculty evaluations and student self-evaluations. In addition to the faculty evaluation of student learning in a given course, program or learning contract, there may appear a one-to-two page detailed description of the program or individual learning contract. Evergreen faculty provide a narrative appraisal of each student's work and progress. These evaluations describe in detail what the student did in the program or contract, what the student was attempting to do, where the student's area of concentration lay, and how well the student succeeded. The faculty evaluation of student work also lists a set of Course Equivalencies, that divides the credit earned into its constitutent parts, and assigns them rough titles to aid other schools or future employers in translating the credit earned into

117

approximations of standard courses. Each student also writes a self-evaluation, describing their work in their own words, explaining what was most important to them, what was unimportant, and why. This self-evaluation also offers evidence of their comprehension and provides details about their progress and success in the program. Students also have the right and responsibility to evaluate their faculty sponsors and seminar leaders; this student evaluation becomes part of the transcript record as well. The final week of every quarter is "Evaluation Week." It corresponds to the conventional exam week, except that is devoted entirely to writing and discussing student and faculty evaluations.

Five Ways to Configure a Program of Study at Evergreen

At Evergreen students have the special opportunity to study one topic at a time from a variety of perspectives, in the arts, humanities, natural, and social sciences. They include **Coordinated Studies, Group Contracts, Individual Learning Contracts, Internships, and Part-Time Studies.** Every program is different, but there are similarities. All Evergreen programs involve a great deal of contact in small groups.

In **Coordinated Study** programs, faculty sign a faculty covenant among themselves regarding the way in which they will conduct the program. In many programs, a second agreement of covenant, **student covenant**, is prepared by the faculty outlining student rights and responsibilities. Among the more important points such covenants usually cover is how credit will be awarded, in what amounts, for what activities, and just what a student must accomplish in order to earn credit. Students in **Coordinated Studies** attend a general lecture with all members of the program and two or more faculty, small group discussions or seminars, field or laboratory sessions, and individual sessions with their seminar faculty. In **Coordinated Study** programs, faculty decide the amount of

credit that can be earned, the subject areas in which it can be earned, and the requirements for earning it.

A Group Contract involves concentrated work with one faculty member and only up to 20 students.

An Individual Learning Contract is a plan of specified activities and consultations, agreed upon verbally and in writing, between the student and a faculty sponsor.

About two-thirds of Evergreen's graduates participate in some form of Internship, either in the context of a fully integrated academic program or separately.

It is possible to pursue Part-Time Studies in a variety of ways: full-time programs with half-time options, specially designed half-time programs, individual learning contracts, and internships. These programs meet at times convenient for working students, usually in the evenings or on weekends.

The Structure of the Evergreen Curriculum

The curriculum spans more than 40 different subjects and is divided into twelve areas of concentration, including Annual Programs, Core Programs, eight Specialty Areas, and two professional programs, one in Teacher Certification and the other a graduate program leading to a Masters degree in Public Administration.

Annual Programs change each year according to the needs and interests of students and faculty or the expertise of a visiting faculty member.

Designed to develop intellectual skills, self confidence, and maturity to meet the demands of college, Core Programs usually lasts a full year and are the academic introduction to the college for entering students although it is not uncommon for juniors and seniors to enroll, whatever their ages or past academic

experience. Their content is broadly interdisciplinary and structured to provide extensive work on oral and written communication skills, close student-faculty interaction, skills in problem solving and critical thinking. Each of the Core Programs listed below is an integrated unit that combines a number of different activities (seminars, individual conferences, lectures, laboratories).

Specialty Areas

The interdisiplinary Specialty Areas include Environmental and Marine Studies, European and American Studies, Expressive Arts, halth and Human Development, Management and the Public Interest, Northwest Native American Studies, Political Economy, Scientific Knowledge and Inquiry. The following list provides samples of the disciplines usually included in Specialty Areas:

Environmental and Marine Studies: biology, geology, planning, natural history, geography, social science, agriculture, physics, mathematics, oceanography, ecology, anthropology, community studies. This specialty area emphasizes field ecology and natural history, marine studies, environmental design and planning, and small-scale agriculture.

European and American Studies: literature, history, philosophy, art history, philosophy social science, political science. These programs study the historical and political trends, artistic and literary documents, social patterns, symbols, religious beliefs and ideological convictions that comprise the way we think and understand our past and future as well. Study in this specialty area relies on the disciplines of literature, history and philosophy, and the disciplinary sub-areas of art history, social and economic history, cultural history, aesthetics and literary theory, and third world studies.

Expressive Arts: visual art, sculpture, drawing, painting, dance, theater, art history, communications, film, video media, music, arts management, aesthetics, crafts. This specialty area offers a variety of basic courses. Students have the opportunity to study creative work in one or more arts, including visual art, film photography, creative writing, drawing sculpture, ceramics, painting, art history, communications, printmaking, music, dance, video media, arts management, aesthetics, costume design, and theater crafts.

Programs in Health and Human Development examine the interaction between biology and psychology of humans and consider a variety of questions and issues within a broad social, ethical, economic, and political context. This specialty area is the focal point for study in human biology, sociology, anthropology, counseling, psychology, history, nutrition, statistics, economics, political sciences, philosophy, health, human services, and education.

The Management and the Public Interest program develops students' abilities to work in the private and/or public sector in public service and leadership roles. Work in this area includes study in management, accounting, marketing, economics, finance, history, philosophy, statistics, political science.

Northwest Native American Studies serve two different student groups: (1) Native American students who are interested in preserving and enhancing their unique cultural heritage and who are developing strategies for self-determination in the world today, and (2) non-Native Americans interested in traditional Native American cultures and values, anthropology, ethnohistory, and the dynamics of cultural change. This specialty area serves as preparation for advanced studies in history, sociology, political science, anthropology, and education.

Political Economy examines the interrelationships of social, cultural, economic, and political phenomena as aspects of an organic whole. Topics of

study include the historical development of the United States and other industrialized nations; the problems of underdeveloped societies in their relations with industrialized societies; the historical contexts in which theories of political economy are developed and applied; and the application of theory to contemporary problems. Study in this specialty area draws upon the sociology, anthropology, literature and law.

Faculty and students in **Scientific Knowledge and Inquiry** examine certain analytical methods and ways of thinking--logical, philosophical, mathematical, and experimental. They study them both for their own sake (in fields such as mathematics, logic, computer science, and analytical philosophy) and as tools for the natural sciences. The traditional nature sciences, particularly physics, chemistry, and biology, fall within this specialty area, but students study them in a broad cultural framework which emphasizes the sciences in relation to the rest of civilization.

Summary

The Evergreen State curriculum presents a unique structure and organization wherein study is organized into programs, courses and learning contracts around a unified theme or issue of interest to the individual student. The major common intellectual experience for students occurs within the programs and specialty areas. Here, students and faculty come together to study interdisciplinary problems, issues and comparative methodology. Content within programs may vary according to the students' needs and interests, as is so reflected in the faculty member's assignment of course equivalencies of credit to the individual student's program description within the narrative evaluation of learning. The assignment of course equivalency credits, thus, is relative to the faculty member's understanding of disciplines and disciplinary specialities,

rather than by some uniform and centralized system of determination. The search for an empirical basis for classifying learning within the Evergreen programs is perhaps more problematic than the classification difficulties encountered by Boysen (1979), Fosdick (1984), Tenopir (1985), or Dressel and DeLisle (1969).

Prerequisities do not appear to play the overt role in constraining student choice found in more traditional curricular structures. Nevertheless, certain programs within each specialty area foresage other programs, courses and individual learning within and between specific disciplines.

Study within programs which may span several quarter terms of enrollment reduces the total number of entities reported on a students transcript. While the number of traditional quarter hour courses on a students transcript may range from 40 to 60 credits, the number of Evergreen programs, courses and learning contracts may be fewer due to the fact that several may span more than one term of enrollment.

123

VI. Qualitative Cluster Analysis of

Evergreen State College Sample

Overview

This section reports the use of the qualitative cluster analytic procedure
to analyze the Evergreen State sample. The findings from the analysis of the
sample is presented. The objects of these analyses are the programs which
constitute the enrollment patterns of students in the Evergreen State sample.
The demographic profile of these Evergreen State students was presented in
Section 3. In Section 4, the distribution of GRE and SAT scores was presented.
Also in that Section, some basic information on the distribution of courses on
the students' transcripts was also presented. The subject of the overall
research is the programs in which students enrolled, not the students
themselves. The criterion variables in the research are the GRE item-type
residuals. The distribution of those residuals among the programs is described
below.

Evergreen State College Sample

There were 269 programs listed on the 46 transcripts of the students in the
Evergreen State sample, indicating that, on average, each of these students had
enrolled in 6 programs as part of the baccalaureate degree studies.

Figure 6-1. The distribution of GRE item-type residuals for 190 unduplicated
programs in Evergreen State College Sample

GRE Item-types	Number of Items	Max Value	Min Value	Score Range	Residual Means	Std Error of Mean	Std Deviation
Analogy	18	1.00	0	1	.5800	.0332	.4583
Sentence Completion	14	1.00	0	1	.5846	.0335	.4618
Reading Comprehension	22	1.00	0	1	.5382	.0340	.4682
Antonyms	22	1.00	0	1	.5790	.0333	.4590
Quantitative Comparison	30	1.00	0	1	.4816	.0339	.4673
Regular Mathematics	20	1.00	0	1	.5350	.0335	.4624
Data Interpretation	10	1.00	0	1	.5869	.0336	.4632
Analytical Reasoning	38	1.00	0	1	.4330	.0338	.4660
Logical Reasoning	12	1.00	0	1	.3992	.0329	.4534
GRE Item-types:							
Minimum	10	1	0	1	.3992	)329	.4534
Maximum	38	1	0	1	.5869	.0340	.4682
Mean	21	1	0	1	.5172	.0336	.4627
Total	186				4.1375		

Figure 6-2. The distribution of GRE item-type residuals for 43 Evergreen programs
used in the Qualitative Cluster Analytic Procedure (Procedure 2).

GRE Item-types	Number of Items	Max Value	Min Value	Score Range	Residual Means	Std Error of Mean	Std Deviation
Analogy	18	1.00	0	1	.5163	.0479	.3143
Sentence Completion	14	1.00	0	1	.5132	.0521	.3418
Reading Comprehension	22	1.00	0	1	.5643	.0514	.3372
Antonyms	22	1.00	0	1	.6283	.0443	.2905
Quantitative Comparison	30	1.00	0	1	.5000	.0504	.3303
Regular Mathematics	20	1.00	0	1	.5267	.0462	.3032
Data Interpretation	10	1.00	0	1	.6165	.5073	.3326
Analytical Reasoning	38	1.00	0	1	.5180	.0533	.3497
Logical Reasoning	12	1.00	0	1	.4382	.0463	.3036
GRE Item-types:							
Minimum	10	1	0	1	.4382	.0443	.2905
Maximum	38	1	0	1	.6283	.5073	.3497
Mean	21	1	0	1	.5382	.1064	.3236
Total	186				4.3052		

Fig 6-3. Comparison of standard deviation

Standard deviation of residuals

```
 .5
.45
 .4
.35
 .3
.25
 .2
     SC  ANA  RD  ANT  QC  RM  BI  ARE  LR
```

190

43

GRE item-type residuals
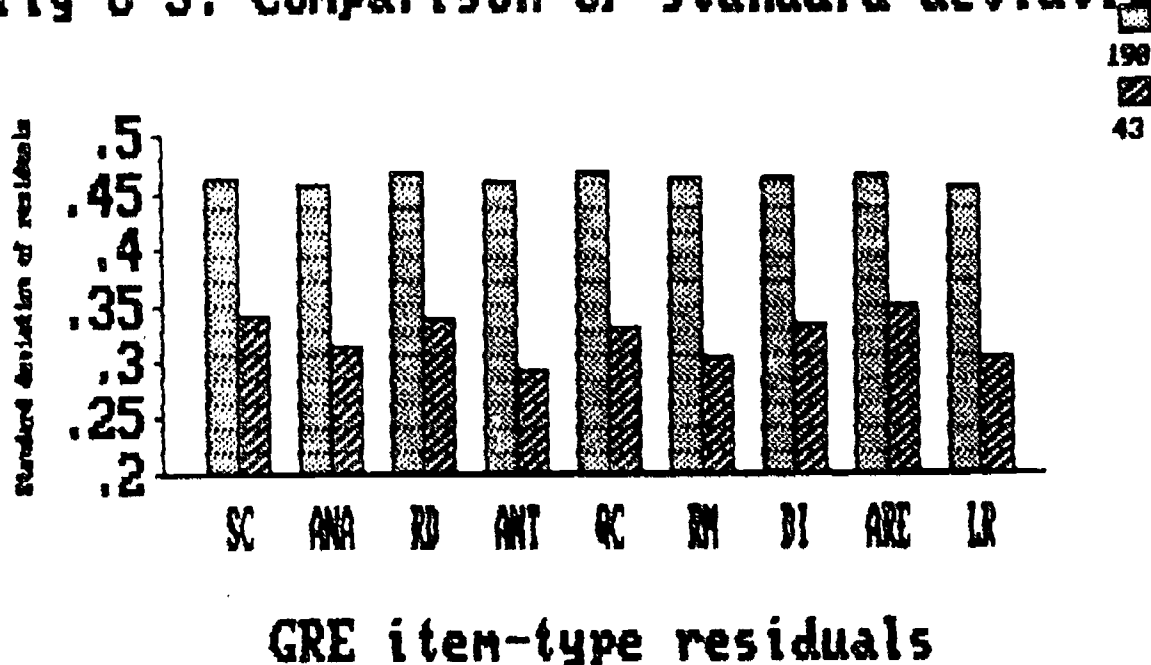
For each program taken by the Evergreen State Sample, the mean of GRE item-type residuals for students enrolled in the course was calculated. As means for unduplicated programs and for programs enrolling 2 or more students were calculated, certain trends emerged. Examination of Figures 6-1 to 6-3 show that as student data were aggregated according to courses enrolling 2 or more students, the standard deviation of these means was considerably smaller than those for all unduplicated programs. This trend suggested that there are relationships between the coursework taken and the student residual scores on the tests.

Cluster Analysis of Evergreen State Sample

Creating the raw data matrix and the resemblance matrix

Using the mean residuals of the Evergreen State Sample and the 43 programs found on 2 or more of the Evergreen student transcripts, a raw data matrix was created. The data matrix consisted of 43 columns and 9 rows (43 x 9). The rows represented the criterion variables: the 9 GRE item-type scores. The columns represented those programs enrolling 2 or more students. Thus, each cell value of the matrix was a mean GRE item-type residual score for each program.

A resemblance matrix was created next to describe how closely each program resembles the other 42 programs according to the criterion variables: the student residuals. To calculate the resemblance matrix, the correlation coefficient was selected as a similarity measure. The correlation coefficient was the Pearson product-moment correlation coefficient. Thus, this coefficient assesses a pattern similarity of any two programs explained in terms of the 9 GRE item-type residuals.

The resemblance matrix produced in this step consisted of 43 rows and 43 columns (43 x 43), in which each cell value theoretically ranged from -1.00 to 1.00. The calculation of the resemblance matrix was done using the SPSSx PROX-IMITY program. The program provides 37 different proximity measures. Ten of them are for quantitative data and the remainder are for binary or qualitative data. This program can directly produce distance, dissimilarity, or similarity matrices as text files for a small to moderate number of cases and variables, which can be directly used for other SPSSx procedures or for other statistical programs. BMDP P1M can also be used to calculate the correlation resemblance matrix and save it as a text file.

127

Selection of the clustering method

The method selected for this analysis was the average linkage method (UPGMA). The appropriateness of the Average Linkage Method was discussed in the previous project reports (Ratcliff, 1988a, 1988b). Contrary to reports by Romesburg (1984), UPGMA is now available in SPSSx, SAS, and BMDP. All the three statistical packages provide UPGMA as an option of the cluster program and use an identical method to compute distances between clusters. The original dendrogram of the Evergreen State Sample programs produced by SPSS-X is presented in Figure 6-4.

Programs were classified into 8 patterns according to a hierarchical cluster structure. In fact, the choice to present the data in 8 clusters is arbitrary. Any number of clusters can be identified depending on the hierarchical cluster structure produced; this structure remains constant regardless of the number of clusters used to form coursework patterns. A procedure for selecting the optimum number of clusters and for validating the resulting patterns will be described in greater detail in a subsequent section on the discriminant analysis of the coursework patterns in the Evergreen State Sample.

Using an 8-cluster solution to the quantitative cluster analysis, the largest number of programs are found in Clusters #3, #4, and #5 each with 8 programs. The smallest clusters are the 7th and 8th clusters each with 1 program. Overall, the differentiation between clusters is attributable to the number of criterion variables used in the analysis and also to the choice of those variables. The cluster analysis and subsequent discriminant analysis suggested that student residuals on GRE item-types are strong, reliable and robust measures in differentiating student general learned abilities.

The hierarchical cluster structure is presented in the dendrogram summary of Figure 6-5. For concise visual presentation, the complex sub-structures of each of the clusters were omitted from the dendrogram. The dendrogram displays the clusters being combined and the distances between the clusters at each successive step, suggesting that the 8-cluster solution examined is appropriate and interpretable. Cluster analyses using smaller and larger numbers of cluster groupings provided comparably high levels of correct classification, as determined by subsequent discriminant analyses. However, as the resemblance index increases (Euclidean distance betweencourses), more distant courses are joined into larger and larger clusters. A 5-cluster solution, for example, provides a high degree of aggregation which may prove to have a high degree of predictive validity but a low level of utility in differentiating coursework by item-type.

129

Figure 6-4. SPSS-X Dendrogram.

RESCALED DISTANCE CLUSTER COMBINE

```
                                          0         5        10        15        20        25
                                          +---------+---------+---------+---------+---------+
Development: The Aim of Education
Foundation of Human Expression
Human Health & Behavior
Theatre & Character
Society & the Computer
Molecule to Organism
Directing Original Theatre
Studio Project
Perspectives on American Culture
Performance Media
Intro to Performing Arts
Commedia Del 'Arte
The Classical World
International Political Economy
Political Economy & Social Change
The Paradox of Progress
Art History of the World
Images & Ideas of Human Ecology
Great Books
Political Ecology
Classical American Drama
Anthropology of Development
Legislative Internship
The Marine Environment
Puget Sound Vertebrate Research
Thinking Straight
The Experience of Fiction
Management & the Public Interest
Human Condition
Gender Issues
Political Economy & Social Change
Great Questions & Great Books
Political Economy & Social Change
Public Ecology
Political Ecology
Matter & Motion
Physical System
Human Development
U.S. History
American Political Economy
Techniques of Visual Anthropology
Inventing America
The Moving Image
```

Figure 6-5. Dendrogram of hierarchical cluster structure for Evergreen State
===

Average Euclidean distance ->

```
            0       5       10      15      20      25
            +-------+-------+-------+-------+-------+-------+

Cluster #4  ------------------------------------+
                                                |
                                                |
                                        +-------+
                                        |       |
                                        |       |
Cluster #6  ----------------------------+       +-+
                                                | |
                                                | |
Cluster #7  ------------------------------------+ |
                                                  |
                                                  |
Cluster #2  --------------------------------------+-+
                                                  | |
                                                  | |
                                                  | |
                                                  | |
Cluster #3  ------------------------------+-------+ |
                                          |         |
                                          |         |
                                      +-+ |         |
                                      | | |         |
                                      | |           |
Cluster #8  --------------------------+ |           |
                                        |           |
                                        |           |
                                        |           |
Cluster #1  ----------------------------+           |
                                        | |         |
                                        | |         |
                                        | |         |
                                        +-----+     |
                                              |     |
                                              |     |
Cluster #5  ----------------------------------+     |
```

131

- 116 -
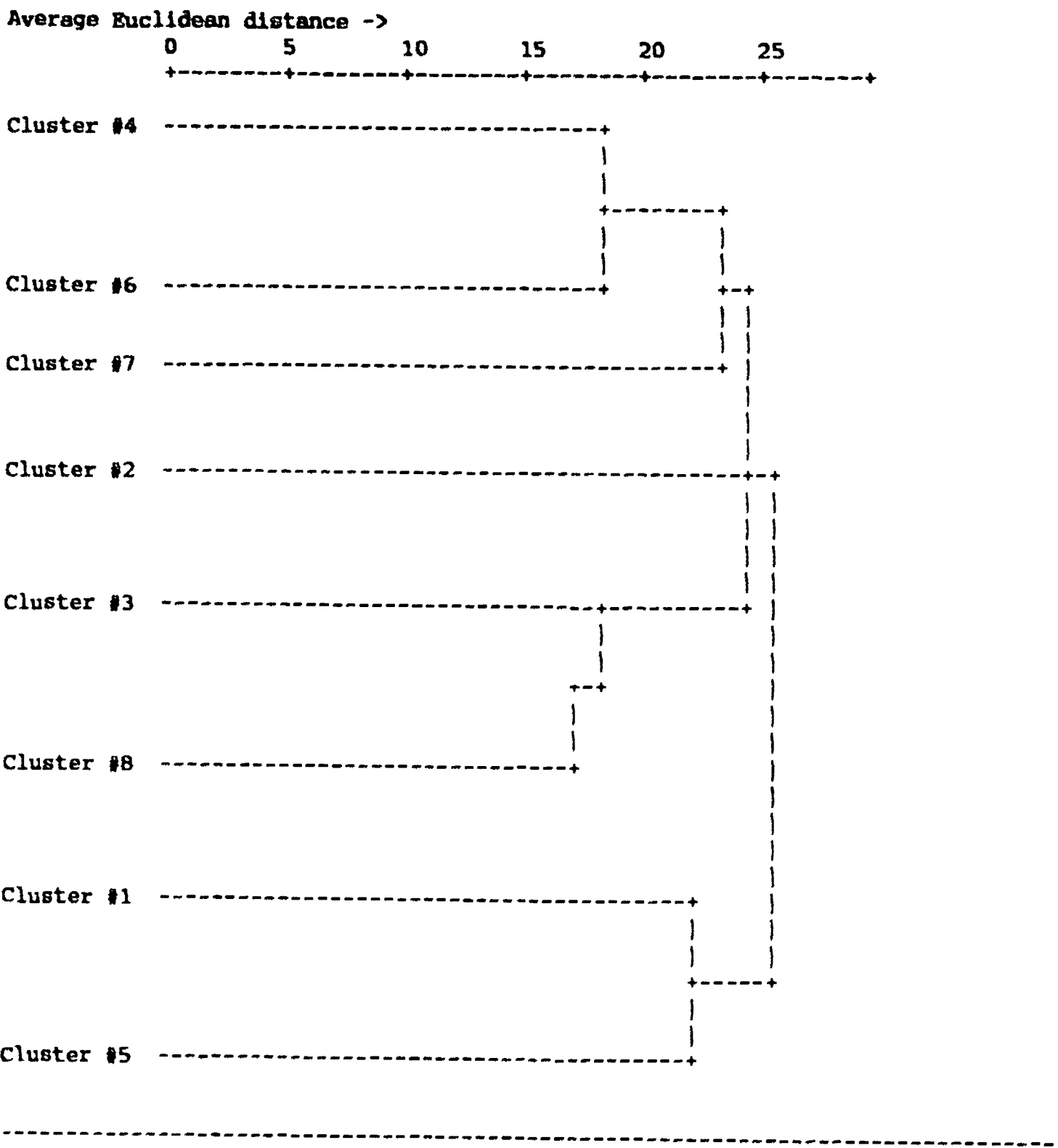
Figure 6-6. Programs within 8-cluster solution: Evergreen

==

Cluster #1 n = 7		Cluster #2 n = 6	
Public Ecology	3098P	Images & Ideas of Human Ecology	4099G
Great Questions & Great Books	4092P	The Paradox of Progress	4130P
Human Condition	5153P	Great Books	5143P
Political Ecology	5203P	Art History of the World	5664C
Political Economy & Social Change	5204P	International Political Economy	6065P
Political Economy & Social Change	7115P	Political Economy & Social Change	6095P
Gender Issues	7735C		

==

Cluster #3 n = 8		Cluster #4 n = 8	
Political Ecology	4137P	Development: The Aim of Education	5127P
Thinking Straight	4157P	Human Health & Behavior	5147P
Anthropology of Development	5018G	Society & the Computer	5215P
The Marine Environment	5173G	Studio Project	6111P
Classical American Drama	5379G	Molecule to Organism	7090P
Puget Sound Vertebrate Research	6131G	Theatre & Character	7095G
Legislative Internship	6775C	Foundation of Human Expression	84084
The Experience of Fiction	7035G	Directing Original Theatre	8994C

==

Cluster #5 n = 8		Cluster #6 n = 4	
Human Development	5154P	Perspectives on American Culture	5196P
Inventing America	6071P	Intro to Performing Arts	6070P
Matter & Motion	6077P	Performance Media	6090P
U.S. History	6647P	Commedia Del 'Arte	7087G
The Moving Image	7092P		
Physical System	7112G		
20th Century American Economy	7122G		
Techniques of Visual Anthropology	7146G		

==

Cluster #7 n = 1		Cluster #8 n = 1	
The Classical World	6013P	Management & the Public Interest	6075P

==

Observations about the Clusters

At this point in the analysis, it is difficult to describe which dimensions
of student general learned ability each cluster represents. However, it seems
to be clear that certain patterns of program enrollment may contribute to
student general learned ability in a way significantly different from the other

programs. Supporting this is a more detailed examination of subset programs of each cluster.

133

Discriminant analysis of program patterns

In examining the dendrogram of the Evergreen State Sample, a logical
question arises as to which number of clusters or pattern groupings provides the
best explanation of the relationship between student item-type score gains and
program patterns. Separate discriminant analyses of different numbers of
cluster groupings were be performed in order to determine the number of
groupings that optimizes the proportion of programs correctly classified. Four
different cluster solutions provided comparably high levels of correct
classification:

     5 cluster solution : 100.0% of courses correctly classified
     8 cluster solution : 100.0% of courses correctly classified
    11 cluster solution : 97.67% of courses correctly classified
    13 cluster solution : 97.67% of courses correctly classified

While these cluster solutions produced comparable classification results,
the different grouping evidenced differing effectiveness in identifying
relationships between mean item-type residuals and program patterns. The
8-cluster solution was used in this report.

The discriminant analysis was conducted using the DISCRIMINANT program in
SPSSx in the following manner. Using the program item-type attributes as
independent variables and the cluster group membership as the dependent
variables, discriminant functions were applied to the data. The discriminant
functions and the program item-type mean residual scores were used to see how
correctly the discriminant function identifies each cluster group. The
resulting percentage of correct predictions serves as a secondary validation of
the cluster solution (Bradfield and Orloci, 1975; Green and Vascotto, 1978;
Romesburg, 1984).

Figure 6-7. Discriminant analysis of the 8-cluster solution for Evergreen State
==

Actual Cluster	No. of Cases	Predicted Group Membership							
		Gr 1	Gr 2	Gr 3	Gr 4	Gr 5	Gr 6	Gr 7	Gr 8
Group 1	7	7 100.0%	0 .0%	0 .0%	0 .0%	0 .0%	0 .0%	0 .0%	0 .0%
Group 2	6	0 .0%	6 100.0%	0 .0%	0 .0%	0 .0%	0 .0%	0 .0%	0 .0%
Group 3	8	0 .0%	0 .0%	8 100.0%	0 .0%	0 .0%	0 .0%	0 .0%	0 .0%
Group 4	8	0 .0%	0 .0%	0 .0%	8 100.0%	0 .0%	0 .0%	0 .0%	0 .0%
Group 5	8	0 .0%	0 .0%	0 .0%	0 .0%	8 100.0%	0 .0%	0 .0%	0 .0%
Group 6	4	0 .0%	0 .0%	0 .0%	0 .0%	0 .0%	4 100.0%	0 .0%	0 .0%
Group 7	1	0 .0%	0 .0%	0 .0%	0 .0%	0 .0%	0 .0%	1 100.0%	0 .0%
Group 8	1	0 .0%	0 .0%	0 .0%	0 .0%	0 .0%	0 .0%	0 .0%	1 100.0%

Percent of "Grouped" Clusters correctly classified: 100.0%

==

Correlations of item-types and discriminant functions

The discriminant analysis of Evergreen State Sample provided secondary validation that 100% of the classification of programs was correctly predicted by the cluster analysis (See Figure 6-7). The discriminant analysis is a secondary validation, since it is based on the same sample of transcripts and test scores. Stated simply, all 10 programs most frequently taken by Evergreen State Sample students were correctly classified according to their mean residual GRE scores. While the cluster analysis produces program patterns according to criteria of general student learning, additional steps are needed (1) to determine which programs were correctly classified and (2) to ascertain which item-type scores contributed to any given program pattern.

Using the BREAKDOWN procedure in the DISCRIMINANT program of SPSS-X (Norusis, 1985), programs which were incorrectly classified or which may be classified within another pattern are identified. When a set of data is taxonomized by multiple and independent criteria, it is reasonable to expect that the first groupings are generally the most homogeneous and the later groupings are the most heterogeneous. However, the 8-cluster solution contained no misclassified programs.

To compute the contribution of each mean item-type residual score to the discriminant functions, the correlation coefficients between mean residual scores and discriminant functions were examined. Figure 6-8 shows the pooled within-group correlations for the 8-cluster solution of Evergreen State Sample coursework.

Figure 6-8. Pooled within-group correlation between mean item-type residual
          scores and discriminant functions for Evergreen State Sample.
===

Mean Residual Item-Type	Func 1	Func 2	Func 3	Func 4	Func 5	Func 6	Func 7
Analytic Reasoning	-.3262	.1570	.0218	.0503	.1714	-.2238	.1451
Quant. Comparisons	-.0143	.0741	.2132	.6073*	-.6054*	-.1569	.0296
Regular Mathematics	-.0079	.1345	-.3193	.4611	.2469	.0472	.6303*
Reading Comprehension	-.0274	.5273*	-.2427	.2769	-.1342	.0200	.3863
Antonyms	.0203	.2247	.3004	.1116	.1443	.8563*	-.1219
Sentence Completion	.2398	-.0387	-.1661	-.3390	.0616	.1448	.3825
Analogies	.0347	-.2690	.2080	-.0196	.3351	-.4363	.7149*
Data Interpretation	-.1108	-.4973*	-.1593	.5576*	-.0726	.2203	.4644
Logical Reasoning	-.2731	.0261	.0835	-.1628	-.4600	.0894	.8194*

Note. Mean residual item-type correlations ordered by size within function.
     Mean item-type residuals with large coefficients are presented in bold,
     grouped together and are marked with asterisks (*).
===

Correlations of coursework clusters and discriminant functions

Figure 6-8 summarizes relationships between GRE item-type residuals and

discriminant functions:

Function 1 was not strongly correlated with the item-types;

Function 2 was positively correlated to Reading Comprehension (r=.53),
          and was negatively correlated to Data Interpretation (r=-.50);

Function 3 was not strongly correlated with the item-types;

Function 4 was positively correlated to Quantitative Comparisons (r=.61),
          and was positively correlated to Data Interpretation (r=.56);

Function 5 was negatively correlated to Quantitative Comparisons (r=-.61);

Function 6 was positively correlated to Antonyms (r=.86); and

Function 7 was positively correlated to Analogies (r=.71),
          was positively correlated to Logical Reasoning (r=.82); and
          was positively correlated to Regular Mathematics (r=.63).

The pooled within-group correlations established relationships between the discriminant functions and the GRE item-type residuals. Each discriminant function explains a certain proportion of the variation in residual scores.

Discriminant functions with strong explanatory power, "good discriminant functions," have large between-cluster variability and low within-cluster variability (Haggerty, 1975; Norusis, 1985). The eigenvalues of Figure 6-9 present the ratio of between-group to within-group sums of squares of the residuals. Large eigenvalues are associated with the di_criminant functions that most contribute to explaining variability in GRE item-type scores.

Wilk's Lambda is the ratio of the within-group sum of squares to the total sum of the squares. It represents the proportion of the total variance in the discriminant function values not explained by differences among cluster groups. Wilk's Lambda serves as a test of the null hypothesis that there is no difference in the mean residuals of a program cluster mean and the mean residual scores of the programs in the total sample. Thus, the eigenvalues and canonical correlations indicate the extent to which each discriminant function contributes to our understanding of the variability in program mean residuals. Lambda test the null of the differential coursework hypothesis for each discriminant function. Functions 1 to 5 explained 97 percent of score variance. Function 1 explained 36 percent of variance; Function 2 explained 31 percent of variance; and Function 3 explained 15 percent of variance. Significant difference occurred by cluster means and sample means affirming the differential coursework hypothesis.

Figure 6-9. Canonical discriminant functions: Evergreen State Sample.

Functio	Eigen-Value	Percent of Variance	Cumulative Percent	Canonical Correlation	Wilk's Lambda	Degrees Freedom	Signifi-cance
0					.0008	63	.0000
1	6.2093	36.08%	36.08%	.9281	.0060	88	.0000
2	5.3791	31.26%	67.34%	.9183	.0380	35	.0000
3	2.6595	15.45%	82.79%	.8525	.1390	24	.0000
4	1.6615	9.65%	92.44%	.7901	.3699	15	.0042
5	.8675	5.04%	97.48%	.6816	.6908	8	.1346
6	.3958	2.30%	99.78%	.5325	.9642	3	.7482
7	.0371	.22%	100.00%	.1891			

Figure 6-10. Canonical discriminant functions evaluated at group means.

Cluster	Func 1	Func 2	Func 3	Func 4	Func 5	Func 6	Func 7	Minimum	Maximum	Average
Cluster #1	3.5266	1.8904	-2.1356	.7707	1.5366	.1729	.6583	-2.1356	3.5266	.9171
Cluster #2	-.3106	-.6608	.2533	-2.3296	-.1069	2.1798	1.6574	-2.3296	2.1798	.0975
Cluster #3	1.0567	-1.9236	-.2615	-.6754	-1.9446	-.2563	-.6870	-1.9446	1.0567	-.6702
Cluster #4	-1.8663	-1.1344	-.5620	1.1447	.4436	-.2510	.0692	-1.8663	1.1447	-.3080
Cluster #5	-1.5191	2.4977	1.7632	1.2046	-.5974	-1.5495	.3444	-1.9191	2.4977	.2491
Cluster #6	-.0556	-1.9697	1.3515	-.8377	1.4717	-.2225	-.3362	-1.9697	1.4717	-.0855
Cluster #7	-1.2975	1.0561	1.7318	-2.2085	.8328	.8327	-2.4515	-2.4515	1.7318	-.2149
Cluster #8	.5272	2.0374	-1.2263	.7509	-.0465	2.2220	-.5683	-1.2263	2.2220	.5281
Minimum	-1.9191	-1.9697	-2.1356	-2.3296	-1.9446	-1.5495	-2.4515			
Maximum	3.5266	2.4977	1.7632	1.2046	1.5366	2.2220	1.6574			
Average	-.0423	.2241	.1143	-.2725	.1986	.3910	-.1642			

Interpreting the coursework clusters for the 8-clus'er solution

Figure 6-10 shows the program cluster means (group centroids) for each discriminant function. Clusters with positive or negative means greater than 1.0 were selected for further analysis.

Coursework Cluster #1 had high positive means on Functions 1, 2, and 5, and a high negative mean on Function 3. Functions 1 and 3 were not strongly correlated with the item-types. Function 2 was positively correlated to Reading Comprehension ($r=.53$) and was negatively correlated to Data Interpretation ($r=-.50$). Function 5 was negatively correlated to Quantitative Comparisons ($r=-.61$). Therefore, students who enrolled in this program pattern represented in Cluster #1 improved in Reading Comprehension but declined on Data Interpretation and Quantitative Comparison item-types.

Cluster #2 had a high negative mean on Function 4. This function was positively correlated to Quantitative Comparisons ($r=.61$) and Data Interpretation ($r=.56$). Students enrolling in this cluster of programs declined in Quantitative Comparisons and Data Interpretation.

Cluster #3 had high negative means on Functions 2 and 5. This evidence suggested that students enrolling in Cluster #3 programs showed gains on Data Interpretation and Quantitative Comparisons and declined on Reading Comprehension.

Cluster #4 had a high negative mean on Function 2 and a high positive mean on Function 4. Students enrolling in Cluster #4 programs improved on Data Interpretation and Quantitative Comparisons but declined in Reading Comprehension.

Cluster #5 had high positive means on Functions 2 and 4. Students enrolled in this set of programs gained in ability on Reading Comprehension and Quantitative Comparisons. The results for the item-type of Data Interpretation were inconclusive.

Cluster #6 had a high positive mean on Function 5 and a high negative mean on Function 2. Students enrolling in Cluster #6 programs improved in Data Interpretation but declined in Quantitative Comparisons and in Reading

Comprehension.

Clusters #7 and #8 consisted of one program each and no further analysis was conducted on these clusters.

Figures 6-11a--6-'1f portray the program clusters and the mean residual item-type with which they were found to be associated. It should be cautioned that the association was established at the cluster level. No direct causal link is intimated between student enrollment in any one given program and scores on the GRE. Furthermore, at this point, one cannot say why students who enrolled in these programs had higher residual scores. The cluster serves to hypothesize relationships between program patterns and the general learned abilities measures by the item-types of the GRE. One can say that students who enrolled in specific patterns of programs tended to improve on specific item-types within the GRE, while others who enrolled in different program patterns did not tend to show such improvement. This evidence affirms the hypothesis that student gains in general learned abilities are associated, positively and negatively, with the programs in which they enrolled. Further analysis is required to determine the nature of these associations.

Figure 6-11a. Cluster 1: High positive mean residuals on Reading Compre-
hension (RD); high negative mean residuals on Data Interpre-
tation (DI) and Quantitative Comparisons (QC).

CLUSTER	DEPT NUM	DESCRIPTION
1	3098P	Public Ecology
1	4092P	Great Questions & Great Books
1	5153P	Human Condition
1	5203P	Political Ecology
1	5204P	Political Economy & Social Change
1	7115P	Political Economy & Social Change
1	7735C	Gender Issues in Modern Political Philosophy

Figure 6-11b. Cluster 2: High negative mean residuals on Data Inter-
            pretation (DI) and Quantitative Comparisons (QC).

```
CLUSTER    DEPT  NUM   DESCRIPTION

   2             4099G  Images & Ideas of the Human Ecology
   2             4130P  The Paradox of Progress
   2             5143P  Great Books
   2             5664C  Art History of the World
   2             6065P  International Political Economy
   2             6095P  Political Economy & Social Change
```


Figure 6-11c. Cluster 3: High positive mean residuals on Quantitative
            Comparisons (QC) and Data Interpretation (DI); high negative
            mean residuals on Reading Comprehension (RD).

```
CLUSTER  DEPT    NUM   DESCRIPTION

   3             4137P  Political Ecology
   3             4157P  Thinking Straight
   3             5018G  Anthropology of Development: General America
   3             5173G  The Marine Environment
   3             5379G  Readings in Classical American Drama
   3             6101G  Puget Sound Vertebrate Research
   3             6775C  Legislative Internship
   3             7035G  The Experience of Fiction
```


Figure 6-11d. Cluster 4: High positive mean residuals on Quantitative Comparisons
            (QC) and Data Interpretation (DI); high negative mean residuals
            on Reading Comprehension (RD).

```
CLUSTER   DEPT   NUM   DESCRIPTION

   4             5127P  Development: The Aim of Education
   4             5147P  Human Health & Behavior
   4             5215P  Society & the Computer
   4             6111P  Studio Project
   4             7090P  Molecule to Organism
   4             7095G  Narrative Life: Theatre & Character
   4             84084  Foundation of Human Expression
   4             8994C  Directing Original Theatre
```

Figure 6-11e. Cluster 5: High positive mean residuals on Reading Comprehension
(RD) and Data Interpretation (DI).

```
=============================================================================
CLUSTER    DEPT    NUM    DESCRIPTION
=============================================================================
    5              5154P  Human Development
    5              6071P  Inventing America
    5              6077P  Matter & Motion
    5              6647C  U.S. History: Founding through Civil War
    5              7092P  The Moving Image
    5              7112G  Physical System
    5              7122G  Race/Gender/Class in 20th Century
    5              7146G  Techniques of Visual Anthropology


=============================================================================
```

Figure 6-11f. Cluster 6: High positive mean residuals on Data Interpretation
(DI); high negative mean residuals on Reading Comprehension (RD)
Quantitative Comparisons (QC).

```
=============================================================================
CLUSTER    DEPT    NUM    DESCRIPTION
=============================================================================
    6              5196P  Perspectives on American Culture
    6              6070P  Introduction to Performing Arts
    6              6090P  Political Economy & Social Change
    6              7087G  Commedia Del 'Arte


=============================================================================
```

VII. Summary and Conclusion

This report concentrated on an examination of transcripts and assessment test scores from the combined sample of graduating seniors at Evergreen State College in the 1987-88 and 1988-89 academic years. The main purpose of this project was to determine if enrollment in different patterns of programs were associated with gains in the general learned abilities of undergraduate students. The answer to this question was consistently "yes". Roughly all of the programs analyzed were accurately grouped according to differential effects in the general learned abilities of students. Taking different patterns of programs did lead to different types and levels of development as measured by the nine item-types of the GRE General Test.

Several consistent findings emerged from the analysis of program clusters. First, the development of general learned abilities did not have an exact one-to-one relationship with departmental categories. All quantitative reasoning development did not occur exclusively in Mathematics-related programs, for example. Consequently, simple counts of the number of credits or courses/programs a student has taken in a particular subject may not be a reliable proxy of general learning in the attendant subject area. Quantitative skills may be developed in a variety of interdisciplinary programs as well. Second, the development of general learned abilities was not confined to the lower division. While there is clearly a difference between general education and general learning (as measured by the GRE), general education requirements should be re-examined in light of student improvement in general learned abilities. Coursework or programs taken by students who showed significant gains should be examined, evaluated and incorporated into the general education sequence of the college. Third, beyond the college catalog, there was little

- 129 -

144

formal monitoring and description of the curriculum in terms of general learned abilities at the college-wide or university-wide level. Colleges should regularly monitor the number of credits and courses in their curriculum. Without this baseline data, the extent to which students share a common learning experience at a college cannot be readily determined.

The relationships established through the cluster analytic model were associational, not causal. Once a set of courses or programs has been linked to improvement in a specific learned ability, a targeted investigation can be launched to determine the commonalities of teaching-learning environment, of student and faculty expectations of performance, of the specific abilities of the students who enrolled in the classes. But regardless of what hypotheses are generated about why this coursework/programs are associated with gains in learned abilities, one can state with confidence that students who enrolled in this coursework demonstrated gains on a specific type of learned ability.

Coursework/program patterns with negative means have limited meaning. A negative mean of residuals on a coursework/program cluster does not necessarily mean actual decline in general learning, only decline about the group tested. Regression automatically defines half the residuals as negative. The mean performance of the group tested is the basis upon which the individuals' GRE item-type scores are predicted from the corresponding SAT scores. By definition, half the group falls below the mean, half is above it. Therefore, those with negative means may have gained in general learned abilities. The negative sign shows that their gain fell below the mean of students in the sample. Declines in general learned abilities are relative to the sample group and may or may not represent actual declines in abilities.

Gains in student learned ability may be attributable to learning outside the classroom. Programs in this research are the settings for analysis.

Comparable analysis could be conducted according to student residential groupings. It is known, for example, that on many campuses the academic performance of students in one fraternity or sorority may be consistently above the average for that college, while students living at another Greek residence setting may be well below the campus norms for academic performance. Does living in a fraternity or sorority cause higher or lower academic performance? Not necessarily. These are selective and self-selected residential situations; the relationship with academic performance is associational.

The coursework/program patterns identified in this research include general education, major, minor and electives. No a priori distinction was made according to these categories prior to analysis of the data. A physical education course in Tennis may be associated with improved learning in Regular Mathematics. This does not mean that enrolling in Tennis causes improved abilities in Regular Mathematics. What it does mean is that students who enrolled in this course tended to improve in this learned ability. Why? The cluster analytic model does not tell us why. Its purpose is to sort through the hundreds or thousands of courses in the college curriculum. The model points out those coursework patterns taken by students who improve in general learned abilities.

The cluster analytic model of analysis is admittedly complex. It would be simpler to calculate residuals on one item-type, such as Regular Mathematics, and then to rank order all coursework according to the mean residuals of students in each course. This would give a picture of each course according to one measure of general learning. However, it would not give an idea of the role of that strength of that ability or measure relative to other measures of general learning used in the assessment. The discriminant analysis shows the role of various types and measures of general learning relative to the

coursework that students take in college.

Where does this leave us? First, we acknowledge that the cluster analytic model performed well with the combined sample at Evergreen State College. It sorted and classified programs according to a given set of measures of general learned abilities. Second, some program patterns identified made sense. Some programs involving mathematics were associated with gains in mathematics abilities.

If all student coursework/program was distributed randomly throughout the curriculum, so too would be their test score residuals. A non-random distribution of residuals implies that specific coursework/programs makes a difference in the development of general learned abilities. Only those courses/programs selected by students showing improvement above the mean for their group should have positive mean course residuals. This research affirmed the differential coursework hypothesis. This research also affirmed the person environment fit hypothesis (Pascarella, 1985).

Student need not choose from several hundred or several thousand courses to have an effective education. The existence of a multitude of courses may be a healthy sign. It may show that colleges and universities accurately mirror the explosion and complexity of knowledge in the last decade of the twentieth century. Such complexity need not impair the development of general learned abilities in undergraduates.

Yet, this research suggests that many students do not make wise course selections. At least they don't make smart choices in relation to those general learned abilities tested by the most commonly recognized post-baccalaureate test of general learning. The mean residuals of most courses taken by students in common with other students of the same cohort was near zero. This means that the general effect of coursework on student learning was randomly distributed

- 132 -

across the curriculum. Perhaps only one-third of coursework taken in common (5 or more students) could be found to have a positive relationship with general learning as measured. This rate of return can be improved by developing more discrete arrays of coursework applicable to general education requirements and by organizing and informing the student academic advising process so that students may choose among coursework aligned with their abilities and prior learning. To that end, the cluster analytic model for identifying the differential effect of coursework holds promise to revising and enhancing the conditions for excellence in undergraduate general learning.

It is commonly assumed that the general curriculum leads to gains in general learning and the major, minor and elective curriculum leads to gains in subject area learning. If such were the case, then the positive residuals on measures of general learning should occur with the courses which students hold in common. Also, negative mean course residuals should be primarily distributed among courses which students did not hold in common.

We know a great deal about what colleges say should be the goals and standards for a baccalaureate degree. The DCP research suggests that much future research is needed to determine what curricular patterns and trends consistently produce the gains in general learning that institutions seek to impart to their students. The challenge of understanding the specific impact of coursework on the learning of students has just begun.

BIBLIOGRAPHY

Adelman, C. (1989). The validity of institutional mission. Paper presented at the Annual Meeting of the Association for the Study of Higher Education, Atlanta, GA, November 2-5, 1989.

Adelman, C. (1988). Opening remarks. "Linking Student Outcomes and Curricular Reform: A New Research Model", a forum presentation at the annual meeting of the American Association for Higher Education (AAHE), Washington, DC, March 9-13, 1988.

Adelman, C. (Ed.). (1986). Assessment in American higher education: Issues and contexts. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Adelman, C. (1985). The standardized test scores of college graduates, 1964-1982. Washington: Study Group on the Conditions of Excellence in American Higher Education, National Institute of Education.

Altmeyer, R. (1966). Education in the arts and sciences: Divergent paths. Unpublished paper, Carnegie Institute of Technology.

Anderberg, M.R. (1973). Cluster analysis for applications. New York: Academic Press.

Arjmand, O., Benbow, C.P., Lorenz, F.O. (Unpublished). Attitude and parental encouragement patterns among intellectually talented males and females. Manuscript, Iowa State University.

Astin, A.W. (1965). Who goes where to college? In A. Astin (ed.), Who goes where to college? Chicago: Science Research Associates, 1965.

Astin, A.W. (1968, August 16). Undergraduate achievement and institutional excellence. Science, 161, 661-668.

Astin, A.W. (1970a). The methodology of research on college impact. Part I. Sociology of Education, 43, 223-254.

Astin, A.W. (1970b). The methodology of research on college impact. Part II. Sociology of Education, 43, 437-450.

Astin, A.W. (1977). Four critical years: Effects of college on beliefs, attitudes. and knowledge. San Francisco: Jossey-Bass.

Astin, A.W. (1985). Achieving educational excellence: A critical assessment of priorities and practices in higher education. San Francisco: Jossey-Bass.

Astin, A.W., King, M.R., & Richardson, G.T. (1979). The American college freshman: National norms for 1979. Los Angeles: Graduate School of Education, University of California.

Ayres, Q.W., & Bennett, R.W. (1983, October). University characteristics and student achievement. Journal of Higher Education, 54, 516-532.

Beeken, L.A. (1982). The general education component of the curriculum through transcript analysis at three Virginia Community Colleges. (Doctoral dissertation, Virginia Polytechnic Institute and State University, 1982).

Belzer, J. et al. (1971). Information science: Analysis and development. Journal of the American Society for Information Science, 22, 193-223.

Belzer, J. et al. (1975). Curriculum information science: Four year progress report. Journal of the American Society for Information Science, 26, 17 31.

Benbow, C.P., & Stanley, J.C. (1980). Sex differences in mathematical ability: Fact or artifact? Science, 210, 1262-1264.

Benbow, C.P., & Stanley, J.C. (1982). Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. American Educational Research Journal, 19, 598-622.

Benbow, C.P., & Stanley, J.C. (1983). Differential course-taking hypothesis revisited. American Educational Research Journal, 20 (4), 469-573.

Bennett, M.H. (1975). A study of the relationship between curricula taken and the critical thinking abilities of high school seniors and University of Illinois freshmen, sophomores, and seniors, majoring in elementary education. (Doctoral dissertation, University of Illinois at Urbana-Champaign, 1975).

Bereiter, C., & Freedman, M. (1962). Fields of study and the people in them. In N. Sanford (Ed.), The American college. New York: John Wiley.

Bergquist, W.H. (1979). Given everything ... success? Evaluation report on the Institutional Development Program, University of Bridgeport, Bridgeport, Conn.

Bergquist, W.H., Gould, R.A., & Greenberg, E.M. (1981). Designing undergraduate education. San Francisco: Jossey-Bass.

Biglan, A. (1973a). The characteristics of subject matter in different academic areas. Journal of Applied Psychology. 57 (3) 195-203.

Biglan, A. (1973b). Relationships between subject matter characteristics and the structure and output of university departments. Journal of Applied Psychology. 57 (3) 204-213.

Birnbaum, R. (1983). Maintaining diversity in higher education. San Francisco: Jossey-Bass.

Blackburn, R., Armstrong, E., Conrad, C. Didham, J., & McKune, T. (1976). Changing practices in undergraduate education. Berkeley, CA: Carnegie Council for Policy Studies in Higher Education.

Blashfield, R.K. (1976). Questionnaire on cluster analysis software. Classification Society Bulletin, 3, 25-42.

150

Bloom, B.S. (1981). All our children learning: A primer for parents, teachers, and other educators. New York: McGraw-Hill.

Boli, J., & Katchaourian, H. (1984a). Analyzing academic records for informed administration: The Stanford curriculum study. Palo Alto: Stanford University, Office of Undergraduate Research. (ERIC Document Reproduction Service, ED 248 787).

Boli, J., & Katchaourian, H. (1984b). Degrees granted and course enrollments by field of study. Second in a series about the Stanford curriculum study. Palo Alto: Stanford University, Office of Undergraduate Research. ED 248 788.

Boli, J., & Katchaourian, H. (1984c). Sex differences in study characteristics and performance. Second in a series about the Stanford curriculum study. Palo Alto: Stanford University, Office of Undergraduate Research. ED 248 789.

Boli, J., & Katchaourian, H. (1984d). Grades, grading standards, and academic awards. Fourth in a series about the Stanford curriculum study. Palo Alto: Stanford University, Office of Undergraduate Research. ED 248 790.

Borg, W.R., & Gall, M.D. (1983). Educational research: An introduction (4th ed). New York: Longman.

Bowen, H.R. (1977). Investment in learning. San Francisco: Jossey-Bass.

Boyer, C.M., & Ahlgren, A. (1981). Visceral priorities: Roots of confusion in liberal education, Journal of Higher Education, 52, 173-181.

Boyer, C.M., & Ahlgren, A. (1982). 'Visceral priorities' in liberal education: An empirical assessment. Journal of Higher Education, 53, 207-215.

Boyer, C.M., & Ahlgren, A. (1987). Assessing undergraduates' patterns of credit distributuion: Amount and specialization. Journal of Higher Education, 58, 430-442.

Buchanan, J.T. (1982). Discrete and dynamic decision analysis. New York: John Wiley.

Bunn, D.W. (1984). Applied decision analysis. New York: McGraw-Hill.

Burkhalter, B.B., & Schaer, B.B. (1984-1985). The effect of cognitive style and cognitive learning in a nontraditional educational setting. Educational Research Quarterly, 9 (4), 12-18.

Carnegie Council on Policy Studies in Higher Education. (1976). A classification of institutions of higher education. (Rev. Ed.) Berkeley, CA: Carnegie Council for the Advancement of Teaching.

Carnegie Foundation for the Advancement of Teaching, The. (1979). Missions of the college curriculum. San Francisco: Jossey-Bass.

Carnegie Foundation's classifications of more than 3,300 institutions of higher education. (1987, July 8). Chronicle of Higher Education, pp. 25-30.

Center for the Improvement of Undergraduate Education. (1974). Yellow pages of undergraduate innovations. Ithaca, NY: Cornell University Press.

Chacon-Duque, F.J. (1985). Building academic quality in distance higher education. University Park: Pennsylvania State University.

Chickering, A. (1971). Education and identity. San Francisco: Jossey-Bass.

Churchman, C.W. (1971). The design of inquiring systems: Basic concepts of systems and organizations. New York: Basic Books.

A Classification of Secondary School Courses. Project Summary Report. (1982, July). Evaluation Technologies, Inc., Arlington, Va. (ERIC Document Reproduction Service No. ED 225 304).

Cohen, A.M. & Brawer, F.B. (Eds.). (1975). The humanities in two-year colleges: A review of the students. Los Angeles: Center for the Study of Community Colleges. (ERIC Document Reproduction Service, ED 108 727).

Conrad, C.F. (1978). The undergraduate curriculum: A guide to innovation and reform. Boulder, CO: Westview Press.

Conrad, C.F. (1986, February). A brief summary of three reports: AAC, NEH, NIE. Unpublished manuscript, Association for the Study of Higher Education, San Antonio, Texas.

Conrad, C.F. et al. (1985). ASHE reader on academic programs in colleges and universities. New York: Ginn.

Conrad, L., Trisman, D., & Miller, R. (eds.). (1977). Graduate record examination technical service manual. Princeton, NJ: Educational Testing Service.

Coyne, J.C., & Lazarus, R.S. (1980). Cognitive style, stress perception, and copy. In I.L. Kutash & L.B. Schlesinger (Eds.), Handbook on stress and anxiety: Contemporary knowledge, theory, and treatment. (pp. 144-158).

Cronbach, L.J., & Gleser, G.C. (1953). Assessing similarity between profiles. Psychological Bulletin, 50, 456-473.

Cross, K.P. (1976). Beyond the open door. San Francisco: Jossey-Bass.

Davis, J. (1964). Great aspirations: The graduate school plans of American college seniors. Chicago: Aldine.

Dillman, D.A. (1978). Mail and telephone surveys: The total design method. New York: John Wiley & Sons.

Drees, L.A. (1982). The Biglan model: An augmentation. (Doctoral dissertation, University of Nebraska, July 1982).

Dressel, P.L., & DeLisle, F.H. (1969). Undergraduate curriculum trends. Washington, D.C.: ACE

Dumont, R., & Troelstrup, R. (1981). Measures and predictors of educational growth with four years of college. Research in Higher Education, 14, 31-47.

Elton, C. & Rose, H. (1967). Significance of personality in vocational choice of college women. Journal of Counseling Psychology, 14, 293-298.

Ennis, R.J. & Millman, J. (1971). Cornell critical thinking test manual. Champaign-Urbana: Critical Thinking Project, The University of Illinois.

Ewens, T. (1978). Analyzing the impact of competence-based approaches on liberal education, in G. Grant & associates (eds.), On competence (pp. 160-198). San Francisco: Jossey-Bass.

Feldman, K. (1969). Studying the impacts of college on students. Sociology of Education, 42, 207-237.

Feldman, K. (1974). Escape from the doll's house. New York: McGraw-Hill.

Ford, N. (1985). Styles and strategies of processing information: Implications for professional education. Education for Information, 3 (2), 155-32.

Fosdick, Howard. (1984). Trends in information science education. Special Libraries, 75(4), 292-302.

Friedlander, J. (1979). The humanities curriculum in the community junior college. Community Junior College Research Quarterly, 3 (4), 297-309.

Fuhrmann, B., & Grasha, A. (1983). A practical handbook for college teachers. New York: Little, Brown.

Gaff, J.G. (1983). General education today: A critical analysis of controversies, practices and reforms. San Francisco: Jossey-Bass.

Gardiner, J. et al. (1986). Instructional Resources in Higher Education. 2nd ed. Stillwater, OK: Oklahoma State University, in press.

Giddings, W.G. (1985). A study of the performance, progress, and degree achievement of Iowa community college transfer students at Iowa's state universities. (Doctoral dissertation: Iowa State University, October 1985).

Gillespie, P.P. & Cameron, K. M. (1986). The teaching of acting in American colleges and universities, 1920-1960. Communication Education, 35(4), 362-371.

Goldman, R.D., Schmidt, D.E., Hewitt, B.N., & Fisher, R. (1974). Grading practices in different major fields, American Educational Research Journal, 11, 343-357.

Goldman, R.D. & Warren, R. (1973). Discriminate analysis of study strategies connected with college grade success in different major fields. Journal of Educational Measurement, 10, 39-47.

Goodlad, J.I. (1981). Curriculum development beyond 1980. Educational Evaluation and Policy Analysis, 3 (5), 49-54.

Grace, J.D. (1985, March). The professionalization of higher education. Paper resented at the Annual Meeting of the American Educational Research Association. Chicago, IL. (ERIC Document Reproduction Service No. ED 261 619).

Grace, J.D. & Fife, J.D. (1986, February). A profile of student expectations of graduate programs in higher education. A paper presented at the annual meeting of the Association for the Study of Higher Education, San Antonio, Texas.

Grandy, J. (1984). Profiles of prospective humanities majors: 1975-1983. Princeton: Educational Testing Service. Final Report # OP-20119-83.

Grant, C.A., & Sleeter, C.E. (1986, Summer). Race, class, and gender in education research: An argument for integrative analysis. Review of Educational Research, 56 (2), 195-211.

Grochow, J. (1973). Cognitive style as a factor in the design of interactive decision-support systems. (Doctoral dissertation, Sloan School of Management, Massachusetts Institute of Technology, 1973).

Groetsch, J.A. (1986). Student preferences for particular instructional strategies in a master of arts program and their relationship to Kolb's learning style. (Doctoral dissertation, Saint Louis University, 1986).

Hall, D. (1969, Spring). The impact of peer interaction during an academic role transition. Sociology of Education, 42, 118-140.

Halpern, D.F. (ed). (1987). Student outcomes assessment: What institutions stand to gain. New Directions for Higher Education, No. 59. San Francisco: Jossey-Bass, Fall 1987.

Hanson, G.R. (1988). Critical issues in the assessment of value added in education. In T. Banta (ed.), Implementing Outcomes Assessment: Promise and Perils. New Directions for Institutional Research, No. 59, Fall 1988.

Hartnett, R.T., & Centra, J.A. (1977, September/October). The effects of academic departments on student learning. Journal of Higher Education, 48, 491-507.

Hayes, E. (1974). Environmental press and psychological need as related to academic success of minority group students. Journal of Counseling Psychology, 21, 299-304.

Hefferlin, J.B.L. (1969). Dynamics of academic reform. San Francisco: Jossey-Bass.

Heiss, A. (1973). An inventory of academic innovation and reform. Berkeley, CA: Carnegie Commission on Higher Education.

Helson, H. (1947). Adaptation-level as frame of reference for prediction of psychophysical data. American Journal of Psychology, 60, 1-29.

Hendel, D.D. (1977). Curricular choices and academic success of students enrolled in an elective studies degree program, Research in Higher Education, 7, 257-267.

Hesseldenz, J.S., & Smith, B.G. (1977). Computer-prepared questionnaires and grouping theories: Considerations for mail surveys in academic settings. Research in Higher Education, 6, 85-94.

Holland, J. (1963). Explorations of a theory of vocational choice and achievement II: A four year predictive study. Psychological Reports, 12, 547-594.

Kagan, J. & Kogan, N. (1970). Individual variation in cognitive processes. In Ph.H. Mussan (ed.), Carmichael's manual of child psychology. Volume 1. New York: Wiley.

Karas, S. (1983). A study of institutional change: Faculty reaction to a change from a quarter to a semester calendar. (Doctoral dissertation, Iowa State University, 1983).

Katz, D., Kahn, R.L., & Adams, J.S. (1982). The study of organizations. San Francisco: Jossey-Bass.

Kelley, D.S. (1983). Comparison of student perceptions of the learning environment during an academic calendar change. (Dcctoral dissertation, Iowa State University, 1983).

King, P.; Wood, P. & Mines, R. (1989). Problem structure, tests of critical thinking and disciplinary differences: A study of critical thinking among college and graduate students. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, 1988.

Kirk, S.L. (1986). The assessment of learning styles and cognitive development of adult students in a higher education setting: Implications for teaching, administration, and advisement. (Doctoral dissertation, University of Northern Colorado, 1986).

Kirtz, J. (1966). Cultural diversity and character change at Carnegie Tech. (Doctoral dissertation, John Hopkins University, 1966).

Kolb, D.A. (1976a). Learning style inventory: Technical manual. Boston: McBer and Company.

Kolb, D.A. (1976b). Management and learning processes. California Management Review, 18 (3), 21-31.

Kolb, D.A. (1981). Learning styles and disciplinary differences. In A.W. Chickering and associates (eds.), The modern American college: Responding to the new realities of diverse students and a changing society. San Francisco: Jossey-Bass.

Kolb, D.A. (1984). Experiential learning: Experience as the source of learning and development. Englewood Cliffs, NJ: Prentice-Hall.

Kolb, D.A. (1985). Learning style inventory. Boston: McBer and Company.

Kolb, D.A., & Fry, R. (1975). Toward an applied theory of experiental learning. In C. Cooper (ed.), Theories of Group Processes. New York: Wiley

Kolb, D.A., & Goldman, M. (1973, December). Toward a typology of learning styles and learning environments: An investigation of the impact of learning styles and discipline demands on the academic performance, social adaptation, and careers choices cf M.I.T. seniors. Boston: Massachusetts Institute of Technology, Sloan School of Management, Working Paper.

Kolb, D., Rubin, R., & Schein, E. (1972). The M.I.T. freshman integration research project: A summary report. Unpublished report, Massachusetts Institute of Technology.

Krathwohl, D.R., ed., et al. (1964) Taxonomy of educational objectives: Handbook II, affective domain. New York: D. McKay.

Kuhn, T.S. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Lewin, K. (1936). Principles of topological psychology. New York: McGraw-Hill.

Levine, A. (1978). Handbook on undergraduate curriculum. San Francisco: Jossey-Bass.

Liberman, L.G. (1986). The effect of congruence between learning/teaching styles on student retention at Broward Community College. (Doctoral dissertation, Florida Atlantic University, 1986).

Little, T.L. (1973). The relationship of critical thinking ability to intelligence, personality factors, and academic achievement. (Doctoral dissertation, Memphis State University, 1973).

Loacker, G., Cromwell, L., & O'Brien. (1986). Assessment in higher education to serve the learner. In C. Adelman (ed.). Assessment in American higher education: Issues and contexts. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

LoGuidice, T. (1983, October). Global and international education at Lutheran-sponsored and affiliated colleges: A look at what is and what should be. Paper presented at the Annual Conference of the Association of Lutheran Faculties. (ERIC Document Reproduction Service No. ED 252 134).

Lorr, M. (1983). Cluster analysis for social scientists. San Francisco: Jossey-Bass.

Marshall, J.C., & Merritt, S.L. (1986). Reliability and construct validity of the learning style questionnaire. Educational and Psychological Measurement, 46 (1), 257-62.

Maynard, L.S. (1984). Student perceptions of the learning environment under a quarter system at two community colleges. (Doctoral dissertation, Iowa State University, 1984).

McKeachie, W.J., Pintrich, P.R., Lin, Y.G., and Smith, D.A.F. (1986). Teaching and learning strategies: A review of the research literature. Ann Arbor: NCRIPTAL.

Mehrens, W. A. & Lehmann, I. J. (1973). Measurement and Evaluation in Education and Psychology\. New York: Holt, Rinehart and Winston.

Miller, G.A. (1969). A psychological method to investigate verbal concepts. Journal of Mathematical Psychology, 6, 169-191

Moore, J.E. (1982). Student persceptions of the learning environment under a quarter system. (Doctoral dissertation, Iowa State University, 1982).

Moos, R. (1976). The human context: Environmental determinants of behavior. New York: Wiley.

Nasatir, D. (1963). A contextual analysis of academic failure. School Review, 71, 290-298.

National Institute of Education. Study Group on the Conditions of Excellence in American Higher Education. (1984, October). Involvement in learning: Realizing the potential of American higher education. Washington, DC: Government Printing Office.

Nettles, M.T (1987, March 2). The emergence of college outcome assessments: Prospects for enhancing state colleges and universities. Working draft paper 87-1. Princeton, NJ: Educational Testing Service. (Available from New Jersey State College Governing Boards Associaton, 150 West State Street, Trenton, NJ 08608).

Nettles, M.T., Thoeny, A.R., & Gosman, E.J. (1986). Comparative and predictive analyses of black and white students' college achievement and experiences. Journal of Higher Education, 57 (3), 289-317.

Nichols, R.C. (1964). The effects of various college characteristics on student aptitude test scores. Journal of Educational Psychology, 55, 45-54.

Nickens, J. (1970). The effect of attendance at Florida junior colleges on final performance of baccalaureate degree candidates in selected majors at the Florida State University, College and University, 45 (3), 281-88.

Norusis, M.J. (1985). SPSS-X advanced statistics guide. New York: McGraw-Hill.

Oh, E.K. (1976). A study of instructor grading philosophies and practices at Western Michigan University. (Doctoral dissertation, Western Michigan University, 1976). Dissertation Abstracts International, 36, 7239A-7240A. (University Microfilms No. 76-10 ,555).

Pace, C.R. (1974). The demise of diversity? A comparative profile of eight types of institutions. New York: McGraw-Hill.

Pace, C.R. (1979). Measuring outcomes of college: Fifty years of findings and recommendations for the future. San Francisco: Jossey-Bass.

Pace, C.R., & Baird, L. (1966). Attainment patterns in the environmental press of college subcultures. In T. Newcomb and E. Wilson (eds.), College peer groups, Chicago: Aldine.

Pallas, A.M., & Alexander, K.L. (1983, Summer). Sex differences in quantitative SAT performance: New evidence on the differential coursework hypothesis. American Educational Research Journal, 20 (2), 165-182.

Pascarella, E.T. (1985). College environmental influences on learning and cognitive development: A critical review and synthesis. In J. Smart (ed.), Higher education: Handbook of theory and research, Volume I. New York: Agathon Press, Inc.

Pascarella, E.T. (1987). Some methodological and analytic issues in assessing the influence of college. Paper presented at the joint meeting of the American College Personnel Association and the National Association of Student Personnel Administrators, Chicago.

Paykel, E.S. (1975). Nonmetric grouping: Clusters and cliqques. Psychometrika, 40, 297-313.

Perry, W.G. (1968). Forms of intellectual and ethical development in the college years. New York: Holt, Rinehart and Winston.

Peterson's annual guide to four-year colleges, 1985. (1985). Fifteenth edition. Princeton, NJ.: author.

Pike, G. R., & Phillippi, R. H. Generalizability of the differential coursework methodology: Relationships between self-reported coursework and performance on the ACT-COMP exam. Research in Higher Education, 30 (3), 245-260.

Plovnick, M. (1974). Individual learning styles and the process of career choice in medical students. (Doctoral dissertation, Sloan School of Management, Massachusetts Institute of Technology).

Powers, D.E., Swinton, S.S., & Carlson, A.B. (1977). A factor analytic study of the GRE aptitude test. (GRE Board Professional Report 75-11P). Princeton, NJ: Educational Testing Service.

Prather, J.E. & Smith, G. (1976a, April). Faculty grading patterns. Atlanta: Office of Institutional Planning, Georgia State University.

Prather, J.E. & Smith, G. (1976b, May.) Undergraduate grades by course in relation to student ability levels, programs of study and longitudinal trends. Atlanta, GA: Office of Institutional Research, George State University.

Prather, J.E., Williams, J.E., & Wadley, J.K. (1976). The relationship of major field of study with undergraduate course grades: A multivariate analysis controlling for academic and personal characteristics and longitudinal trends. Atlanta, GA: Georgia State University, Office of Institutional Research, Report OIP-77-3. (ERIC Document Reproduction Service, ED 132 898).

Ral, A. (1956). The psychology of occupations. New York: John Wiley.

Ratcliff, J.L. (1985). Time as a dimension in curricular change. A paper presented at the Forum, "Perceived changes on the learning and teaching environments as a result of academic calander change", annual meeting, American Association of Community and Junior Colleges, San Diego.

Ratcliff, J.L. (1987). The effect of differential coursework patterns on general learned abilities of college students: Application of the model to an historical database of student transcripts. Task #3 Report. U.S. Department of Education, Office of Educational Research and Improvement, Contract No. OERI-R-86-0016, Iowa State University, June 1987. 57p.

Ratcliff, J.L. (1988a). The development of a cluster analytic model for determining the associated effects of coursework patterns on student learning", a paper presented at the annual meeting of the American Educational Research Association (AERA), New Orleans, LA, April 1988.

Ratcliff, J.L. (1988b). Development and testing of a cluster-analytic model for identifying coursework patterns associated with general learned abilities of college students: Progress report #6. U.S. Department of Education, Office of Educational Research and Improvement, Research Division. Contract No. OERI-R-86-0016.

Ratcliff, J.L. (1990a). Development and testing of a cluster-analytic model for identifying coursework patterns associated with general learned abilities of college students: Final report, Stanford University samples #1 & #2 February 1990 U.S. Department of Education, Office of Educational Research and Improvement, Research Division. Contract No. OERI-R-86-0016. University Park, PA: Center for the Study of Higher Education, Pennsylvania State University.

Ratcliff, J.L. (1990b). Development and testing of a cluster-analytic model for identifying coursework patterns associated with general learned abilities of college students: Final report, Ithaca College samples #1 & #2, February 1990 U.S. Department of Education, Office of Educational Research and Improvement, Research Division. Contract No. OERI-R-86-0016. University Park, PA: Center for the Study of Higher Education, Pennsylvania State University.

Rice, D.D. (1973). The relationship of lower division specialization to success in major fields of study at Florida State University. Tallahassee, FL: Unpublished doctoral dissertation, Florida State University, 1973.

Richardson, R.C., & Doucette, D.S. (1980). Persistance, performance and degree achievement of Arizona's community college transfers in Arizona's public universities. Tempe, AZ: Arizona State University, Dept. of Higher and Adult Education. (ERIC Document Reproduction Service, ED 197 785).

Richardson, R.C., & others. (1982). A report of literacy development in community colleges: Technical report. Washington, DC: National Institute of Education. (ERIC Document Reproduction Service No. ED 217 925).

Rock, D.A., Baird, L.L., & Linn, R.L. (1972, January). Interaction between college effects and students' aptitudes. American Education Research Journal, 9, 149-61.

Rock, D.A., Centra, J., & Linn, R.L. (1970, January). Relationships between college characteristics and student achievement. American Educational Research Journal, 7, 109-121.

Romesburg, H.C. (1984). Cluster analysis for researchers. Belmont, CA: Lifelong Learning Publications.

Roueche, J.E., & Snow, J.J. (1977). Overcoming learning problems: A guide to developmental education in college. San Francisco: Jossey-Bass.

Rudolph, F. (1977). Curriculum: A history of the American undergraduate course of study. San Francisco: Jossey-Bass.

Schoenfeldt, L.F. & Bush, D.H. (1975). Patterns of college grades across curricular areas: Some implications for GPA as criterion. American Educational Research Journal, 12, 313-321.

Simpson, E.J. (1972) "The classification of educational objectives in the psychomotor domain." Vol. 3. The psychomotor domain. Washington: Gryphon House.

Skinner, S.B. (1976). Cognitive development: A prerequisite for critical thinking. The Clearinghouse, 49, 373-376.

Sloan, D. (1971). Harmony, chaos, and consensus: The American college curriculum. Teachers College Record, 73, 221-251.

Spaeth, J., & Greeley, A. (1970). Recent alumni and higher education: A survey of college graduates. New York: McGraw-Hill.

Stabel, C. (1973). The impact of a conversational computer system on human problem solving behavior. Unpublished working paper, Sloan School of Management, Massachusetts Institute of Technology.

Stadtman, V.A. (1980). Academic adaptations. San Francisco: Jossey-Bass.

Sternberg, R. (1982). Reasoning, problem solving and intelligence. In R. Sternber (Ed.), Handbook of human intelligence, (pp. 225-307). New York: Cambridge University Press.

Strasmore, M. (1973, June). The strategic function re-evaluated from the organization development perspective. Master's thesis, Sloan School of Management, Massachusetts Institute of Technology.

Swinton, S.S., & Powers, D.E. (1982). A study of the effects of special preparation on GRE analytic scores and item types. (Graduate Record Examination Board Report No. 78-2R). Princeton, NJ: Educational Testing Service.

Taba, H. (1962). Curriculum development. New York: Harcourt, Brace.

Tenopir, C. (1985). Information science education in the United States: Characteristics and curricula. Education for Information, 3(1), 3-28.

Terenzini, P. T. (1989). Assessment with open eyes: Pitfalls in studying student outcomes. Journal of Higher Education. 60 (6), 644-703.

Terman, L., & Oden, M. (1947). The gifted child grows up. Stanford, CA: Stanford University Press.

Toombs, W.; Fairweather, J., Chen, A.; & Amey, M. (1989). Open to View: Practice and purpose in general education 1988. A Final Report to the Exxon Education Foundation, August 1989. University Park, PA: Center for the Study of Higher Education, Pennsylvania State University.

Torrealba, D. (1972). Convergent and divergent learning styles. Unpublished masters thesis, Sloan School of Management, Massachusetts Institute of Technology.

Trow, M. (1962). "Student Cultures and Administrative Action." In R.L.S. Sutherland and others (Eds.) Personality factors on the college campus. Austin, TX: Hogg Foundation for Mental Health.

Tryon, R.C. (1939). Cluster analysis. Ann Arbor: Edwards Brothers.

Tyler, R. (1950). Basic principles of curriculum development. Chicago: University of Chicago Press.

Veysey, L. (1973). Stability and experiment in the American undergraduate curriculum. In C. Kaysen (ed.), Content and context: Essays on college education. New York: McGraw-Hill.

Waggaman, J. S. (1980). Surrogate learning measures: Credit, other units and non-credit. Florida State Department of Education, Tallahassee.; Florida State University.

Warren, J.B. (1975). Alternatives to degrees. In D.W. Vermilya (Ed.), Learner-centered reform: Current issues in higher education 1975. San Francisco: Jossey-Bass.

Warren, J.B. (1978). The measurement of academic competence. Final report, Grant #G007603526. Washington, DC: Fund for the Improvement of Postsecondary Education.

Warren, J.B. (1987, March 10). The assessment of learning in mechanical engineering: First progress report. Contract No. 300-87-0003. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Watkins, D. (1986). Learning process and background characteristics as predictors of tertiary grades. Educational and Psychological Measurement, 46 (1), 199-203.

Watson, G. & Glaser, E.M. (1964). Watson-Glaser critical thinking appraisal manual. New York: Harcourt, Brace & World.

White, K.R. (1979). The need for guidance in a general education program. Journal of General Education, 25 (1), 39-42.

Wilson, K.M. (1974). The contribution of measures of aptitude and achievement in predicting college grades. (Education Testing Service Research Bulletin 74-36). Princeton, NJ: Educational Testing Service.

Wilson, K.M. (1985, February). The relationship of GRE general test item-type part scores to undergraduate grades. Princeton, NJ: Educational Testing Service, Report No. ETS-RR-84-38; GREB-81-22P. ERIC Document Reproduction Service ED 255 544.

Witkin, H. (1976). Cognitive style in academic performance and in teacher-student relations. In S. Messick and associates, Individuality in learning: Implications of cognitive styles and creativity for human development. San Francisco: Jossey-Bass.

Witkin, H.A. et al. (1977). A longitudinal study of the role of cognitive styles in academic evolution during the college years. Princeton, NJ: GRE Board Research Report GREB No. 76-10R.

Woditsch, G.A. (1977). Developing generic skills: A model for competency-based general education. Bowling Green, OH: Bowling Green State University.

Wolfe, D. (1954). America's resources of specialized talent. New York: Harper.

Wood, P.K. (1983). Inquiring systems and problem structure: Implications for cognitive development. Human Development, 26, 249-265.