DOCUMENT RESUME

ED 306 275 TM 013 113

AUTHOR Ratcliff, James L.; And Others
TITLE Determining the Effect of Different Coursework Patterns on General Student Learning at Four Colleges and Universities.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Mar 89
CONTRACT OERI-R-86-0016
NOTE 90p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Ability; *Academic Records; Achievement Gains; *Achievement Tests; Cluster Analysis; College Curriculum; *College Graduates; *Course Selection (Students); Higher Education; Learning Processes; Scores; Standardized Tests; Student Educational Objectives; Student Interests
IDENTIFIERS *Course Taking Patterns; *Different Coursework Patterns; Graduate Record Examinations; Scholastic Aptitude Test

ABSTRACT

Preliminary findings of the Differential Coursework Patterns (DCP) Project are reported. The Project uses assessment test scores and transcripts from samples of graduating college seniors to determine what coursework patterns were related to gains in the general learned abilities of the students. Random samples of graduating seniors have been examined from: (1) a major private research university (n=105); (2) a state university (n=54); (3) a private college (n=146); and (4) a women's liberal arts college (n=62). Precollege Scholastic Aptitude test scores were an indication of the students' entering levels of learning. The Graduate Record Examination general tests were the post college exit measure of general learning. Coursework patterns resulting from a cluster analysis developed for the project did not produce clear distinctions according to academic department or major, but it was concluded that: (1) development of general learned abilities did not have an exact one-to-one relationship with departmental categories; (2) development of general learned abilities was not confined to the lower level of courses; and (3) there was little formal monitoring or description of the curriculum in terms of general learned abilities at an institution-wide level. Standardized test scores and transcripts were useful in the assessment of learned abilities. The relationships between coursework patterns and the general curriculum suggest that new ways are needed to conceptualize general education in college. Thirteen tables, three graphs, and 10 figures present study data. A 186-item list of references is provided. (SLD)

ED306275

# Determining The Effect Of Different Coursework Patterns On General Student Learning At Four Colleges And Universities



James L. Ratcliff

and Associates

Higher Education Section

Iowa State University

Ames, IA 50011

A paper presented at the Annual Meeting,
American Educational Research Association,
San Francisco, California, March 27, 1989

This research was conducted as part of Contract No. OERI-R-86-0016, U.S. Department of Education, Office of Educational Research and Improvement, Research Division. The perspectives and conclusions are those of the author and do not represent the views or policy of the U.S. Department of Education.

DETERMINING THE EFFECT OF DIFFERENT COURSEWORK PATTERNS

ON GENERAL STUDENT LEARNING

AT FOUR COLLEGES AND UNIVERSITIES

James L. Ratcliff

and Associaties

Higher Education Section

Iowa State University

This paper reports preliminary findings of the Differential Coursework Patterns (DCP) Project. These findings may have particular significance and interest to state policymakers, test and assessment organizations, and those at colleges and universities who are responsible for academic policy as it pertains to undergraduate education.

The Differential Coursework Patterns Project is a direct result of the research recommendations of the Study Group on the Conditions of Excellence in Higher Education. Those recommendations appeared in the Group's report, Involvement in Learning: Realizing the Potential of American Higher Education. The purpose of the DCP Project was to use assessment test scores and transcripts from samples of graduating college seniors to determine what coursework patterns were related to gains in the general learned abilities of the students.

We have now examined random samples of graduating seniors who began their undergraduate education at: 1) a major private research university, 2) a doctoral-granting state university, 3) a private comprehensive college, 4) a women's liberal arts college, 5) a public, non-traditional liberal arts college, 6) a comprehensive community college and subsequently transferred to a state university. This report excludes discussion of the public non-traditional college due to the lack of comparability of much of the data. It also excludes discussion of the community college data, since it comprises a sub-sample of one of the other institutions studied. It was not our purpose to compare student performance at these institutions. Rather, it was to develop a way to link what coursework students took to what students learned during their undergraduate education.

What we have found so far, we feel, has profound implications. In this report is important information for state policymakers who may be considering implementation or revision of a state assessment program for higher education. We think this report can provide conservative guidance in what is and what is not a realistic expectation from a student assessment program within a statewide system of higher education.

The preliminary findings of the DCP Project also has some important suggestions for such organizations as the Educational Testing Service, the College Board, the Graduate Record Examination Board, and the American College Testing Service. These organizations provide the major sources of standardized test information to colleges and universities. They are currently developing new instruments in response to the assessment movement in higher education. Our preliminary findings suggest some ways that these organizations can be more helpful through their existing instruments.

We have been examining student transcripts and assessments of general learning. In doing so, we have discovered some interesting and provocative relationships between effective coursework patterns and the overall undergraduate curriculum. These relationships suggest that new ways are needed for conceptualizing general education in college. We propose specific ways to approach reform of the undergraduate curriculum. We also propose specific means for selecting and monitoring the effect of the curriculum on the general learning of students.

After making some bold assertions about what this working paper contains, we need to also urge some caution. In the past two years, we at the DCP Project have developed a model of analysis and have used it to look at one sample of student from each of the six institutional environments listed previously. Our analytic model worked in all circumstances. It did identify the specific coursework associated with gains in general learning among college students. Yet, we need a second group of students from each of these institutional environments to confirm our findings. We are gathering that data now. What we present here are hypothesized relationships. They are provocative. But, they need further research to bear them out. While we are examining a second sample of students from each of the six aforementioned environments, the preliminary findings reported here also have taken us beyond the original scope of our investigation. In short, we have discovered some aspects of student assessment and the college curriculum which merit further exploration. We conclude this report, therefore, with recommendations for further research. In essence, we conclude by challenging each of you -- state policymakers, testing service research analysts, college or university curricular leaders -- to join with us in a dialogue to enhance our ability to understanding what in the formal curriculum contributes to the general learning of undergraduate students.

## The notion of student assessment and some of its attendant problems

There is a simple notion upon which the assessment movement is based:

> Assessment of learning outcomes is one solution to this lop-sided notion of educational quality. Instead of documenting excellence by variables unrelated to learning, like research reputation and size ..., we can assess educational quality more directly by examining what and how much students actually learn....
>
> Using student outcomes assessment, institutions can refocus their priorities so that educational quality is based on what and how much students learn in school. (Halpern, 1987, pp. 5-6).

Give students a test at the end of their education. If they do well on the test, they must have received a quality college education. If they do not, the college wasn't such a good investment. A simple, clear idea of quality, right? Not exactly! The idea is one fraught with problems. Some of the problems have been discussed by others, yet bear repeating. Others come directly from our DCP research. None of the problems are insurmountable, but all needed to be addressed prior to implementing an effective assessment program.

One problem is that students at a single college or university do not experience the same education! Incredible as this statement seems, look at the transcript evidence from four of our six institutional environments:

Figure 1. Description of students and their transcripts.

| Samples | College 1 | College 2 | College 3 | College 4 |
|---|---|---|---|---|
| Number of Students | 105 | 54 | 146 | 62 |
| Percent of graduating seniors | 10% | 15% | 14% | 48% |
| College Calendar | Quarters | Quarters | Semesters | Quarters |
| Total number of courses on transcripts | 5,541 | 2,087 | 6,249 | 2,607 |
| Average number of courses taken | 52.8 | 38.6 | 42.8 | 42.0 |
| Total number of Unduplicated courses on transcripts | 1,445 | 815 | 1,136 | 990 |
| Percent of Duplication | 73.92% | 60.95% | 81.82% | 62.03% |
| Unduplicated courses per 100 students | 1,376 | 1,509 | 778 | 1,597 |
| Total number of Unduplicated courses taken by 5 or more students | 303 | 133 | 405 | 149 |
| Percent of Duplication by 5 or more students | 20.97% | 16.32% | 35.65% | 15.05% |
| Unduplicated courses per 100 students | 289 | 246 | 277 | 240 |

## Problems posed in examining curricular effects on undergraduates

Figure 1 shows the sample student transcripts we examined in the DCP Project. The average number of courses taken by these graduating seniors suggests that from 3,900 to 4,300 courses could be found on a set of 100 student transcripts! When courses which appear on more than one transcript were counted only once (Unduplicated courses), the number of courses per 100 students ranged from 778 to 1,597! Five or more students in these samples had from 15.1 to 35.7 percent of their coursework in common. Students from the same college had very different formal educational experiences.

One purpose of student assessments of general learning is to determine how well the college is doing in developing students' general learned abilities. So, if we test these students (as we did), the first assessment problem we need to recognize is that they did not share much of a common intellectual experience at the college or university they attended. That these students experienced so little of the curriculum in common may be viewed as desired or undesirable. The fact remains, <u>it means little to make generalizations about the quality of education at the institution as a whole when students present such a wide diversity of educational experiences on their transcripts</u>.

## Does diversity mean disarray?

Does this diversity of coursework mean disarray? Not necessarily. Let's use the quarter system as an example: It usually takes about 180 quarter credits to earn a B.A. degree. It is not uncommon to require 90 to 120 of these credits (33 to 50 percent) to meet the general education requirements of the college or university, leaving the balance of credits to be earned in the major, minor and electives. Given the variety of majors, minors and electives at a college or university, a truly random sample of student transcripts should produce less than 50 percent of coursework in common. Thus, at least half of the formal curriculum of these graduating seniors _should_ be diverse. Depending upon the range of choices available to students in meeting their general education requirements, the percent of coursework held in common should be equal to or less than one-third of coursework on the transcripts.

The general education constitutes from one-half to one-third of the student's coursework. College's normally concentrate general education coursework in the lower division of the curriculum. They often formally or informally encourage students to complete the majority of the general education requirements during the first two-years of college. The majority of coursework taken at a community college by those students intending to transfer and earn a bachelors is general education and elective coursework. In as much as half of the students earning bachelors degrees are transfer students, and because so many colleges and universities confine their general education requirements to the first lower division courses, it would seem to make sense to assess general learning of undergraduates at the end of the sophomore year -- the so-called "rising junior" examination.

Assessments conducted at the end of the sophomore year, however, are built upon the assumption that what a college _intends_ to be the impact of its curriculum actually _is_ the impact. If a college sets a sequence of mathematics courses as a general education requirement, does this mean that student actually _learn_ their advanced quantitative skills in these courses? What about the junior-level courses in Statistics, Accounting, Market Research, or Test and Measurement in Education? Don't these courses contribute as well to the general learned quantitative abilities of students? In the DCP Project research, we found this to be the case; there was specific upper division coursework associated with for certain types of general learned abilities. The difference between the general education coursework required by the college and the coursework patterns associated with gains in general learned abilities may be attributable to the difference between the intended and the actual effects of the formal curriculum on general student learning.

If a purpose of an assessment activity is examine what students learned as a result of their college experience, then all the curriculum -- not just the general education curriculum -- should be examined to determine its contribution. If we wanted to improve the writing ability of students graduating from a particular college, do we enhance the freshman composition courses? Do we require more composition courses for graduation? Do we expect writing as a requirement for all courses in the curriculum? Because different students take different coursework, there is no simple way to answer this question. The lower performance in writing may be a result of what the student selected rather than inherent weaknesses in courses with the intent of improving one's writing skills. One of the exciting results of the DCP Project research is that we can identify the coursework patterns of those students who did improve in specific general learned abilities. A major gap in the assessment movement has been the lack of a method for linking the coursework in which student enrolled with gains in specific types of general learned abilities.

Because there has not been a means for linking specific coursework with specific types of learning, there is no means for arriving at the conclusion that a required core of courses or a distributional menu of courses is superior in conveying general learning to college undergraduates. Diversity in the curriculum has not been shown conclusively to have either an affirmative or detrimental effect on students' cognitive development. This diversity in the undergraduate curriculum does have implications for assessing student learning. To assess the effect of the formal curriculum on general student learning, we must first be prepared to identify what that curriculum was. And we can't just do that by taking about the institution as a whole or its general education requirements alone. Student experiences with the curriculum are simply too diverse. Instead, our method of analysis must allow for a natural level of diversity in the curriculum, in student interests and abilities, and in the resulting selection of coursework patterns. Such an allowance for diversity adds to the complexity of the task of linking coursework patterns with gains in student learned abilities, but coping with such complexity and diversity is not insurmountable.

Testing how much students learned as undergraduates without looking at what they took in college is a bit like examining the fuel efficiency of the college motor pool fleet without determining how many dump trucks and a how many subcompact cars are in the fleet. To determine the effects of college on student learning, the unit of analysis has to be the coursework taken, not the college as a whole. Similarly, to assess student learning in college without acknowledging the student's prior learning in high school, elementary school and preschool is to measure the college's selectivity in admission, rather than the gains it contributed to the student's cognitive development.

What constitutes a coursework pattern?

The prevalent way to view the college curriculum is by its intentions, rather than by its results. Given that measuring the effect of the curriculum is problematic, it is not surprising that many studies presume, rather 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, 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 to 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." (Perry, 1968, 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 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 fulfilled 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 descriptions 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 subjects required for certification or licensure in a particular profession, occupation or technical field. Such lists of coursework may be compiled by practitioners and academics of a given discipline or profession to accredit college or university programs. Just as the curriculum of a particular college may represent 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 is 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 pattern of data is a design resulting from "the relation among a set of objects" (Romesburg, 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 is a set of one or more courses found to be similar to one another according to a given set of attributes. In the DCP research, courses whose students demonstrated comparable GRE item-type residuals were grouped together. Thus, for the DCP project, a cluster of courses was a set of courses associated by common effects on the general learned abilities of the students enrolling in them. 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.

What are general learned abilities?

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, 1981). 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).

A common stumbling block in the development of an assessment program is that of what form of test or assessment information to use. It is acknowledged that there is no clear conception of what constitutes general learning, either in the college curriculum or in the various tests and assessment devices. If a college attempts to reach consensus among its constituents on either general education goals or on the "best" measure of general learned abilities, there surely will be heated discussion and the quest may end in irresolution or, worse, abandonment of the assessment initiative. Rathering than searching for the ideal measure of general learning in college, <u>those charged with assessment can better direct their energies toward the selection of a constellation of assessment means and measures which appear to be appropriate criteria for describing one or more dimension of the general learning goals of the college.</u> As will be seen, the DCP Project research design provides a basis for determining the relative extent to which each in a set of measures explains student learning within a specific college environment.

Given the confusion of terms, intents and theoretical frameworks given to explain "general learned abilities", we developed an analytic model for the DCP Project research design that is <u>criterion-referenced</u>. That is, the model is not based on any one notion of what constitutes "general learned abilities". It is not dependent on any one college's curricular structure or its intended goals to promote student cognitive development. The DCP analytic model allows the use of multiple and different measures of student outcomes. It provides a means for weighing those various measures to determine the extent to which they reflect the coursework in which the students actually enrolled. The validity of the assessments (multiple measures) of student learning rests on the validity of the outcome measures selected, not on the DCP analytic model.

The DCP model fulfills the need for multiple measures and criterion-referenced measures of student learning. However, the design described below is not dependent upon any given college curricular structure or organization and, in fact, was tested in six very different higher education institutions. It is in free of bias engendered by specific institutional goals or preconceived notions of the curriculum, such as breath or depth of the general education coursework (Cronbach, 1985, p. 212). Thus, "general learned abilities" -- the term used to denote student cognitive development in the DCP Project -- is defined by the criteria selected to measure general student learning. Such criteria, by necessity, should encompass a variety of commonly recognized areas of knowledge, skill and ability development in undergraduate education.

<u>What constitutes student achievement?</u>

While the expansiveness and diversity of the college curriculum has implications for determining what were the college effects on students, there are also a number of problems associated with measuring "general learned abilities." The first of these is establishing exactly what constitutes growth or gains in such abilities. Simply measuring how graduating seniors perform on a series of tests is not a sufficient basis for generalizations about the importance of college. 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 the GRE, measures are attributable to student learning prior to college.

Nichols (1964) studied pre-college and post-college test scores of 381 National Merit Scholars, using SAT-Verbal and SAT-Math scores as student aptitude measures; these scores correlated strongly with the GRE-Verbal and GRE-Quantitative ($\underline{r}^2=.74$), while student/faculty ratio, library books per student, average ability level of the student body, and affluence of the college were all unrelated to GRE scores. Nichols concluded that "the college a student attends does, indeed, have an effect on his performance on an examination such as the GRE" (p. 9).

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. However, the colleges whose students had the same SAT means did not necessarily have similar GRE means. Rock, Baird and Linn found that "for colleges characterized by similar and relatively higher verbal input, the humanities data do suggest that proportion of faculty with the doctorate, size of budget, and selectivity are related to achievement" (p. 158).

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 their learning prior to college. So we should anticipate strong relationships between well-established measures of general learning, such as the SAT and GRE, because both measure comparable types of learning and a large proportion of that learning occurs prior to the students admission to college.

Unfortunately, we have found some assessment programs have not taken into account student learning prior to college. With a high proportion of general learning occurring prior to college, and with no means to account for prior learning in such assessment programs, colleges run the danger of confusing the achievement of the students with the selectivity of the admissions process. If one college admits higher ability students than a second, the assessment results at the first college will be higher than that at the second simply by virtue of the relative number of "grey cells" students brought to the occasion, not necessarily the learning they acquired during their college years. While Astin (1985) has made this point repeatedly in advancing his concept of "talent development" among college students, it is also important in the use of assessment for program review at or between institutions. Higher education academic policy and resources should be directed by accomplishments, not by the extent to which particular institutions can attract students to enroll.

We have noted in professional meetings and in the literature a general antipathy toward standardized tests as measures of general learned ability. The tests have been criticized for not measuring "critical thinking" and "higher order reasoning skills". Furthermore, test scores as reported have little diagnostic value; their very structure and procedures are intended to be predictive rather than diagnostic. The SAT serves to predict how well a student will perform in college in math and verbal abilities. The GRE General Test offers a forecast of how well a student will perform in graduate school in verbal, quantitative and analytic abilities (the three sub-scores of the test), based on their general learning through their years as undergraduates.

In the DCP Project, we used students' precollege SAT scores as indicators of their entering levels of general learning. We use the General Tests of the Graduate Record Examination (GRE) as the postcollege exit measures of general learning for the samples of graduating seniors.

The Normative Base of the Graduate Record Examination

The strong relationship between the SAT and the GRE is both an asset and a liability. The use of the SAT sub-scores as pre-college measures and the GRE item-types as a post-college 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 pre-college and post-college 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 among students 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 bachelors to graduate study than do those who choose to go to college from high school. Nevertheless, a consequence is that the GRE examinations are normed on a higher ability population than the SATs (Pascarella, 1985). The individuals taking the GRE examinations 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). Thus, Graduate Record Examination can be accurately viewed as a measure of a students' predicted performance in graduate school as well as a measure of that students' general academic accomplishments as an undergraduate. Examination of five institutional samples from the DCP Project revealed the students were roughly comparable to the norm of graduating seniors at each college in terms of SAT scores, and distribution of major across the curriculum. Since the analyses were institution specific and criterion-referenced, the role of the norms upon which the GRE were based did not immediately enter into the analysis.

<u>The Graduate Record Examination as a measure of higher order reasoning or critical thinking ability</u>

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. It is argued that a major function of undergraduate education is the development of higher order reasoning or critical thinking skills and that multiple choice, paper and pencil tests do not measure these skills well Yet, these two concepts elude any concise definitions (Skinner, 1976).

One factor differentiating tests of critical thinking is that of problem structure (Sternberg, 1982; Wood, 1983). Problem structure is the extent to which a problem can be described fully and can be answered rightly or wrongly. Complex social, political or economic problems do not have right or wrong answers and often there very nature is debated. These are ill-defined problem sets. In contrast, problems which can be solved by deductive logic (in the spirit of Sherlock Holmes or Miss Marple) possess a high degree of certainty and correctness; they are well-structured problems (Churchman, 1971).

Two popular measures of critical thinking are the Cornell Critical Thinking Test (CCTT) and the Watson-Glaser Critical Thinking Appraisal (WGCTA). Each measures student's ability to solve structured problems (Ennis & Millman, 1971; Watson & Glaser, 1964). Each has been shown to have strong correlations with the ACT, SAT and GRE examinations (Little, 1973; Bennett, 1975/1976). In a recent study, for example, King, Wood & Mines (1988) found that the WGCTA correlated with the ACT at $\underline{r}^2=.59$ and the CCTT correlated with the ACT at $\underline{r}^2=.62$. If sixty percent of the variance in scores on a college-level critical thinking test is explained by a traditional, standardized measure of student achievement, what is the critical thinking test really measuring? The difference between what the standardized measures of general learning measure and what is measured in several of the tests of critical thinking may not be fundamentally different.

Students in the DCP Project samples showed significant improvement in Analytic Reasoning (ARE) and Logical Reasoning (LR) abilities. To what extent do these item-types represent critical thinking or high order reasoning skills? Further investigation is warranted to determine the nature of the knowledge and cognitive abilities required to answer these questions. To what extent do they, for example, suffice a college's question to measure the development of critical thinking abilities in its students?

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 (and women) undergraduates (Astin, 1968; Wondel, 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 the item-type scores on either the GRE or SAT as multiple measures of general learning have not been widely explored (Adelman, 1988).

Research in the DCP Project suggests that the item-types of these tests can be used effectively to describe specific general learned abilities across a broad and representative spectrum of graduating seniors. These seniors come from selective and open-admissions institutions of higher education. While high ability students who demonstrated significant improvement in one or more general learned abilities did enroll in different coursework patterns from students of low entering ability, there was little evidence to suggest that the GRE test failed to differentiate between the pre- and post-college achievement of high ability students. There were not significant instances where the GRE item-type score predicted by the student's SAT score exceeded the actual perfect score for that item-type. In no case did a individual student's 9 predicted GRE scores exceed the 9 possible GRE item-type scores.

Development of the DCP model of analysis

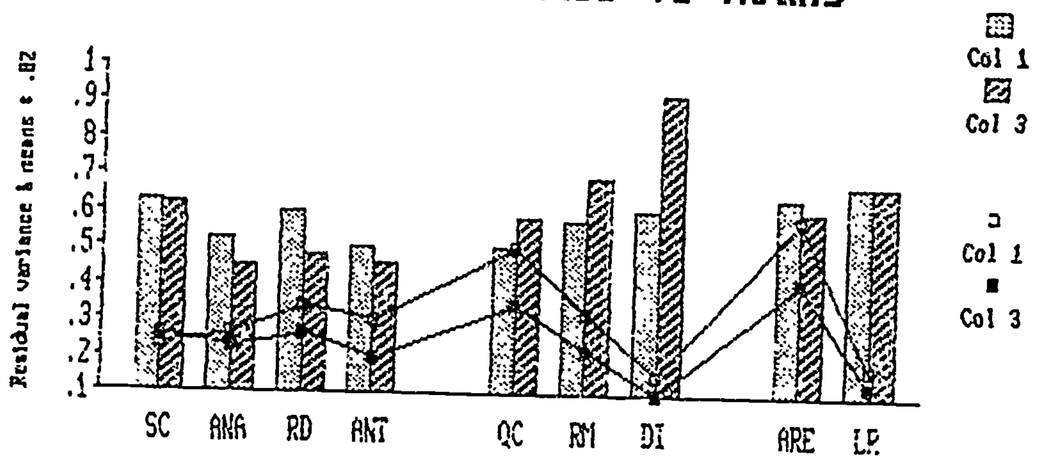
The GRE and SAT tests were used in the development of the DCP Project as pre-college and post-college achievement measures. However, the analytic model developed is not dependent upon these sets of measures. The model could be employed using any set of correlated pre-tests and post-tests. What is needed for the model to function effectively is multiple measures of student learning wherein pre-college student achievement is accounted for and controlled. For pragmatic and economic reasons, the cluster analytic model was developed using SAT and GRE tests, respectively, as pre-test and post-test measures of the general learned abilities of baccalaureate candidates for graduation. 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 measures of student learning. Therefore, the cluster analytic model will be described in terms of 9 measures of student general learned ability. There was prior research to suggest 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. 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 by field of study. Results suggested that the GRE item-type scores did differentiate undergraduate GPA by field of study. This and other studies (Powers, Swinton & Carlson, 1977; Swinton & Powers, 1982; Wilson, 1974) indicate that the 9 item-type subparts of the GRE test measure different and somewhat unique general learned abilities.

The GRE General Test consists 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 (post-college measure) and the GRE scores predicted by the students' corresponding SAT scores (pre-college measure) were used to gauge general learned abilities attributable to the students' undergraduate education. For the purposes of development and testing of the cluster analytic model, 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. Hereafter, these 9 item-type residuals will be referred to as <u>student score gains</u>. Thus in an economical and practical way, these student score gains represent a set of multiple measures of general learned abilities which account for and control the effect of pre-college student achievement on post-college student outcomes.

The advantages of using residual scores rather than mean scores in the design of assessments of undergraduate learning is illustrated in the following comparison of the performance of seniors at two colleges in the DCP project:

Fig 2. Residuals vs Means



Discussion of merits of score residuals over simple score means in determining gains in general learned abilities

Comparison of the mean GRE item-type scores and the GRE residual scores (once the effect of the SAT is removed) illustrate the differences between controlling and not controlling for incoming student ability. Consider the following data from the DCP research: College 1 has extremely selective admissions. Both the verbal and quantitative SAT scores are quite high at College 1. College 3, while maintaining selective admissions as well, has students entering with considerably lower SAT quantitative scores. Figure 6 presents the means (lines) and residuals (bars) for each college's seniors.

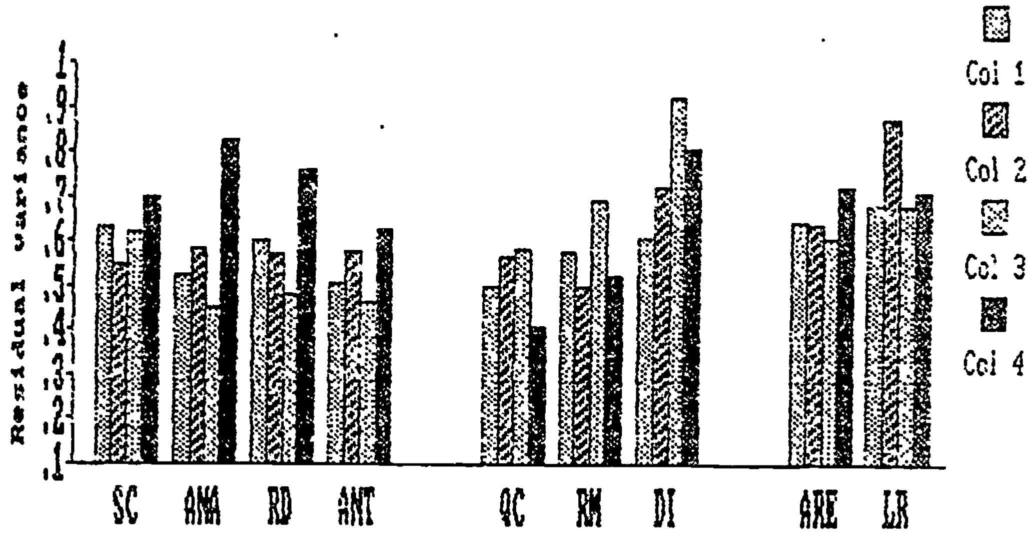
If we examine simple mean score performance on the GRE (and therefore, do not take into account the precollege abilities of the students), we conclude that College 1 seniors evidenced greater skill in most all areas. Their skill in making Quantitative Comparisons, Analytic Reasoning (ARE) and Reading Comprehension (RD) were particularly pronounced. One might falsely conclude from examination of simple mean scores that students in College 1 learn more than those from College 3.

When we examine residual scores (after the effect of SAT scores are removed), we find quite a different story. College 1 students showed significant improvement in verbal abilities, particularly Analogies (ANA) and Reading Comprehension (RD). College 3 students showed significant gains in quantitative abilities, particularly in Data Interpretation skills (DI). Students at both colleges showed significant improvement in all measures of general learned ability, with the residual scores representing from 45 to 92 percent of their learning as measured by the GRE item-types. Residuals tell you more about the performance of students, and they account for their learning prior to college.

Figure 2a compares the mean score residuals for each college sample and each GRE item-type. All colleges demonstrated improvement in the general learning of their students. College 4 showed particularly large gains in Verbal item-types. College 3 exhibited gains in Quantitative abilities, particularly in Data Interpretation. College 2 manifested marked performance improvement in Logical Reasoning. Each college demonstrated a different mix of score gains among the 9 criterion variables. The residual scores provide a basis for determining the extent to which each measure of general learning explains the achievements of a group of college students.

To reconfirm the findings of Wilson (1985), we compared the GRE item-type residuals and GRE sub-score residuals. If the item-types measure discrete types of learning, then a greater proportion of the sub-scores should be explained by the SAT than the item-type scores. In short, the residuals for the GRE item-types should be greater that those for the GRE sub-scores, indicating that less of the variance in item-type scores is explained by the SAT. As the comparison of three college samples in the DCP Project shows (see Figure 3), item-types do evince greater residuals than do sub-scores.

Fig 2a. Comparison of residuals




across college samples

Fig 3. Item-types vs Sub-scores

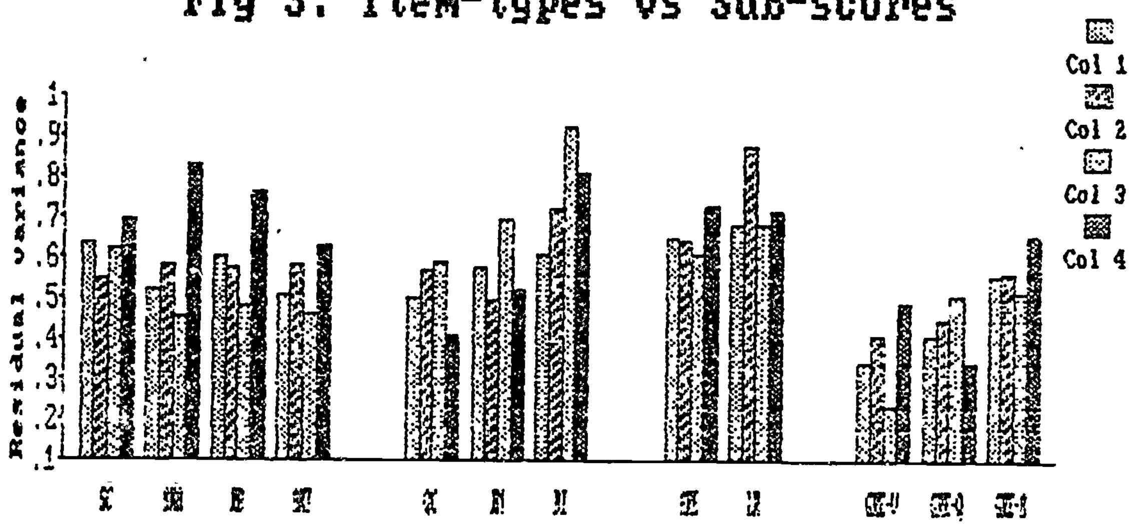



Figure 3 shows that College 4 students showed greater gains from precollege to postcollege measures in verbal abilities. Improvement in student ability to answer Antonym (ANT) and Reading Comprehension (RD) questions were particularly pronounced. College 3 students demonstrated large gains in quantitative skills.

Both College 3 and College 4 showed considerable improvement in Data Interpretation (DI), while College 2 students showed exceptional gains in Logical Reasoning (LR) skills. All four colleges evinced larger gains in specific abilities in individual GRE item-types than in the aggregate GRE sub-scores (GRE-V, GRE-Q & GRE-A).

Two observations are important to note here. First, the item-types scores appeared to measure discrete forms of learning. Secondly, while all colleges showed significant improvement in all areas, item-type residual gains did vary by college. College environments were associated with gains in the general learning of students. Different college environments were linked with different patterns of improvement. And the GRE item-types did differentiate this learning more effectively than did GRE sub-scores.

Given this confirmation of Wilson's findings regarding item-types, it would seem warranted to investigation SAT item-types as well. Wilson asserted that they too were associated with specific types of learning. If such were the case, a more effective set of multiple measures of precollege learning could be used. Unfortunately, these SAT item-type scores were not available to the DCP Project team.

Why should item-types have greater power in describing discrete sets of learning than aggregate sub-scores? This is partly the result of the way the tests were constructed. If the GRE included but one type of mathematics questions, it would be subject to the criticism that it was not a comprehensive measure of students' quantitative skills. So several discrete types of mathematics questions were included in the GRE. Using factor analysis, the creators of the test established the construct validity of each item-type. When item-type scores are aggregated to form a single Quantitative sub-score to predict academic performance, the discrete qualities of each item-type are submerged in the large

## Attributing student score gains to courses

### Reliability and correlation of GRE Item-types

Before removing the effects of the students' SAT scores from their GRE item-type scores, we first tested the reliability of the GRE item-types for each sample of college seniors. We wanted to determine if GRE item-types were reliable measures of student learning for each group of students. Next, we examined the correlation between the GRE item-types and the SAT sub-scores and total scores. We did this to make sure we were using the most appropriate SAT score or sub-score as a control variable. Finally, we conducted a regression GRE item-types on SAT sub-scores to calculate student gain scores (residuals) for each GRE item-type. We wanted to see if Wilson's (1985) assertion the item-types measured discrete types of learning (and hence constituted multiple measures of learning) was affirmed in the sample data from the various sample groups. If such were the case, there should be greater unexplained variance (residuals) in the GRE item-types than in the GRE sub-scores.

Our first question was whether GRE item-types evinced reliability within each sample group. The results of the reliability analysis for each college sample is presented in Figure 4. The reliability of the individual item-types tended to increase with the number of items comprising the given item-type. While those item-types comprised of a small number of items (Data Interpretation and Logical Reasoning) presented low alpha coefficients, overall the GRE item-types demonstrated satisfactory levels of reliability, stability and accuracy of measurement.

Figure 4. The reliability of coefficients of GRE item-types for 3 colleges.

| GRE Item-types | Code | Number of items | Cronbach's Alpha College 1 n = 107 | College 3 n = 62 | College 4 n = 146 |
|---|---|---|---|---|---|
| Analogy | ANA | 18 | .5674 | .6126 | .4414 |
| Sentence Completion | SC | 14 | .6192 | .5980 | .5317 |
| Reading Comprehension | RD | 22 | .7617 | .7013 | .5826 |
| Antonyms | ANT | 22 | .6769 | .8257 | .6054 |
| Quantitative Comparison | QC | 30 | .7814 | .6968 | .7475 |
| Regular Mathematics | RM | 20 | .6650 | .4971 | .7023 |
| Data interpretation | DI | 10 | .5786 | .1817 | .3565 |
| Analytical Reasoning | ARE | 38 | .8055 | .8021 | .7931 |
| Logical Reasoning | LR | 12 | .5034 | .6065 | .3602 |
| GRE Verbal | GRE-V | 76 | .8765 | .8852 | .7943 |
| GRE Quantitative | GRE-Q | 60 | .8686 | .7575 | .8405 |
| GRE Analytic | GRE-A | 50 | .8315 | .8261 | .8091 |

In the DCP Project 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. Determining whether the GRE item-type, Analogies, for example, 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.

As Figure 5, generally the strongest relationships were between the GRE Verbal item-types and the SAT Verbal score. In the case of College #1, GRE Reading Comprehension was more strongly correlated with the SAT Total score than with the SAT Verbal sub-score. Highest correlations occurred between the GRE Quantitative Item-Types and the SAT Math score, and between the GRE Analytic item-types and the SAT Total score. In the instances of Colleges #3 and #4, Data Interpretation was as strongly correlated with the SAT Total Score as with the

Math sub-score. And in the cases of College #3, Analytic Reasoning was about as strongly correlated with the SAT Math sub-score as it was with the SAT Total Score.

The GRE item-type correlations were markedly lower than the GRE sub-score correlations. This meant th  there was greater unexplained variance on the GRE item-types than on the GRE sub-scores. Stated another way, less of the student achievement measured by the SAT was explained the GRE item-type scores than explained the GRE sub-scores. GRE Item-type correlations were frequently 10 percent weaker than their corresponding GRE Sub-score correlation coefficients.

Regressing GRE Verbal Item-Types on SAT Verbal sub-scores, GRE Quantitative Item-Types on SAT Math scores, and GRE Analytic Item-types on SAT Total scores appeared appropriate. In general, the GRE Item-Types and the SAT scores showed significant positive correlation, suggesting that the SAT test as comparable measure of general learned abilities. GRE Item-type correlations were noticeably lower than their corresponding GRE sub-scores, suggesting that individual item-types also measure discrete abilities apart from those of the SAT sub-scores.

Figure 5. The correlation of GRE item-types and SAT scores in 3 colleges

| | | College #1 | | |
|---|---|---|---|---|
| GRE Item-types | Code | SAT Verbal | SAT Math | SAT Total |
| Analogy | ANA | .6956 | .3820 | .6482 |
| Sentence Completion | SC | .6121 | .4130 | .6101 |
| Reading Comprehension | RD | .6371 | .4781 | .6599 |
| Antonyms | ANT | .7080 | .2732 | .6001 |
| Quantitative Comparison | QC | .3139 | .7117 | .5712 |
| Regular Mathematics | RM | .4137 | .6553 | .6067 |
| Data interpretation | DI | .2887 | .6321 | .5137 |
| Analytical Reasoning | ARE | .4524 | .5961 | .6012 |
| Logical Reasoning | LR | .5134 | .4622 | .5715 |

| | | | | |
|---|---|---|---|---|
| GRE Verbal | GRE-V | .8127 | .4658 | .7674 |
| GRE Quantitative | GRE-Q | .3490 | .7657 | .6227 |
| GRE Analytic | GRE-A | .5292 | .6363 | .6723 |

College #3

| GRE Item-types | Code | SAT Verbal | SAT Math | SAT Total |
|---|---|---|---|---|
| Analogy | ANA | .4257 | .1759 | .3445 |
| Sentence Completion | SC | .5555 | .3363 | .5189 |
| Reading Comprehension | RD | .4953 | .3288 | .4815 |
| Antonyms | ANT | .6164 | .2797 | .5151 |
| Quantitative Comparison | QC | .3277 | .7704 | .6779 |
| Regular Mathematics | RM | .2585 | .6941 | .5908 |
| Data interpretation | DI | .2778 | .4466 | .4404 |
| Analytical Reasoning | ARE | .3041 | .5602 | .5285 |
| Logical Reasoning | LR | .3818 | .5164 | .5421 |
| GRE Verbal | GRE-V | .7176 | .3922 | .6431 |
| GRE Quantitative | GRE-Q | .3523 | .8141 | .7196 |
| GRE Analytic | GRE-A | .6119 | .6119 | .5904 |

College #4

| GRE Item-types | Code | SAT Verbal | SAT Math | SAT Total |
|---|---|---|---|---|
| Analogy | ANA | .7457 | .2972 | .6365 |
| Sentence Completion | SC | .6244 | .9700 | .5542 |
| Reading Comprehension | RD | .7359 | .4511 | .7061 |
| Antonyms | ANT | .7399 | .3388 | .6551 |
| Quantitative Comparison | QC | .3216 | .6512 | .5542 |
| Regular Mathematics | RM | .2470 | .5649 | .4603 |
| Data interpretation | DI | .3081 | .3038 | .3598 |
| Analytical Reasoning | ARE | .5032 | .5852 | .6352 |
| Logical Reasoning | LR | .6324 | .3214 | .5769 |
| GRE Verbal | GRE-V | .8714 | .4170 | .7820 |
| GRE Quantitative | GRE-Q | .3041 | .3805 | .3983 |

| | | | | |
|---|---|---|---|---|
| GRE Analytic | GRE-A | .6071 | .5724 | .6955 |
| | | | | |
| Minimum GRE item-type | | .2470 | .1759 | .3445 |
| Maximum GRE item-type | | .7457 | .9700 | .7061 |
| Mean GRE item-type | | .4865 | .4887 | .5615 |
| | | | | |
| Minimum GRE sub-score | | .3041 | .3805 | .3983 |
| Maximum GRE sub-score | | .8714 | .8141 | .7820 |
| Mean GRE sub-score | | .5728 | .5618 | .6546 |

$p < .0001$

The internal validity of GRE item-types can be measured by comparing the inter-correlation coefficients of GRE item-types. In the various DCP Project college sample groups, the inter-correlations between GRE Verbal item-types were relatively stronger than those between other GRE item-type scores. Each GRE sub-score tended to have higher correlations with the GRE item-types constructing the sub-score than with GRE item-types constructing other test sub-scores. The analysis of correlations among GRE item-types shows that the item-types have strong internal validity.

Figure 6 presents the GRE raw score results for item-types and sub-scores for each of four colleges. College 1 students performed very well on the GRE examination. They answered an average 57 of 76 Verbal (GRE-V) questions correctly. They answered 49 of 60 Quantitative (GRE-Q) questions correctly, and 38 of 50 Analytic (GRE-A) questions correctly. This corresponds to the following GRE converted scores: Verbal- 615, Quantitative - 701, Analytic - 688. One or more students attained a perfect score on each item-types except Antonyms (ANT); no student attained a perfect score on more than one item-type.

College 4 students performed well on the GRE examination. They answered an average 47 of 76 Verbal (GRE-V) questions correctly. They answered 33 of 60 Quantitative (GRE-Q) questions correctly, and 27 of 50 Analytic (GRE-A) questions correctly. This corresponds to the following GRE converted scores: Verbal- 522, Quantitative - 472, Analytic - 531. A few students attained perfect scores on the Analogies (ANA) and Antonyms (ANT) item-types; no students attained perfect scores on more than one item-type. To determine the extent to which these students evinced gains over their precollege SAT scores, the GRE raw scores were regressed on the corresponding SAT scores.

Regressing GRE Verbal Item-Types on SAT Verbal sub-scores, GRE Quantitative Item-Types on SAT Math scores, and GRE Analytic Item-types on SAT Total scores appeared appropriate. GRE Item-type correlations were consistently and noticeably lower than their corresponding GRE sub-scores in each college sample. This suggests that individual item-types also may measure discrete abilities apart from those of the SAT sub-scores and/or it may reflects their lower reliability stemming from fewer items.

Figure 6. Distribution of GRE Item-Type mean scores among college samples.

| GRE Item-types | Number of Items | College 1 Mean | College 3 Mean | College 4 Mean |
|---|---|---|---|---|
| Analogies | 18 | 13.11 | 8.45 | 11.47 |
| Sentence Completion | 14 | 11.67 | 10.21 | 9.81 |
| Reading Comprehension | 22 | 16.77 | 11.50 | 13.06 |
| Antonyms | 22 | 15.17 | 10.83 | 12.84 |
| Quantitative Comparison | 30 | 25.07 | 4.32 | 17.40 |
| Regular Mathematics | 20 | 16.33 | 11.13 | 10.84 |
| Data interpretation | 10 | 7.55 | 5.04 | 4.97 |
| Analytical Reasoning | 38 | 29.05 | 20.99 | 20.52 |
| Logical Reasoning | 12 | 8.71 | 6.01 | 6.56 |
| Minimum | 10 | 7.55 | 4.32 | 4.[illegible] |
| Maximum | 38 | 29.05 | 20.99 | [illegible].52 |
| Mean | 21 | 15.94 | 9.83 | [illegible]1.94 |
| p > .0001 | 186 | 143.43 | 88.47 | 107.47 |

Calculation of mean score gains for students

As Figure 6 demonstrates, the students in each of four colleges performed well on the GRE General Examination, answering the questions correctly in approximately 113 of the 186 GRE items. GRE raw scores were varied according to college and varied widely. Very different distinctions 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 7), different college groups had the greatest amount of unexplained variance (highest residuals) on different item-types. Likewise, different college groups had the lowest amount of unexplained variance on divergent item-types. Significant differences in unexplained variance existed among college samples.

The greatest amount of variance in item-type score gains (residuals), including the greatest standard error and standard deviation, was found in the Data Interpretation item-type in the College 3 sample. The lowest amount of variance in score gains was evident in Quantitative Comparisons in College 4. Variance was partially explained by the number of items included in each item-type, but unexplained variance still differed with the college sample. The highest unexplained variance for College 1 and College 2 was in the Logical Reasoning item-type, which consists of 12 items. The highest unexplained variance for College 3 was in Data Interpretation, which consists of 10 items, and the highest unexplained variance in College 4 was in Analogies, which consists of 18 items. The lowest unexplained variance was found in Quantitative Comparisons for Colleges 1 and 4; this item-type contains 30 items. The lowest unexplained variance for College 2 was Regular Mathematics, which contains 20 items, and the lowest unexplained variance in College 3 was in Analogies, encompassing 18 items. The improvement in general learned abilities of students at each of the four colleges was explained to a different extent by each of the item-type residuals. Score gains varied by individual college and by item-type.

Figure 7. Regression of SAT scores on GRE item-type scores

| Dependent Variables: GRE Item-Types | College 1 Sample #1 | College 2 Sample #1 | College 3 Sample #1 | College 4 Sample #1 |
|---|---|---|---|---|
| <u>Sentence Completion</u> | | | | |
| F Value | 61.724 | 111.195 | 38.340 | 64.258 |
| probability>F | .0001 | .0001 | .0001 | .0001 |
| R-Squared (SAT-V explained variance) | .3747 | .4572 | .3899 | .3085 |
| Adj. R-Squared | .3686 | .4531 | .3797 | .3037 |
| Residuals (unexplained variance) | .6314 | .5469 | .6203 | .6963 |
| <u>Analogies</u> | | | | |
| F Value | 96.542 | 96.878 | 75.154 | 31.865 |
| probability>F | .0001 | .0001 | .0001 | .0001 |
| R-Squared (SAT-V explained variance) | .4838 | .4233 | .5561 | .1812 |
| Adj. R-Squared | .4788 | .4189 | .5487 | .1755 |
| Residuals (unexplained variance) | .5212 | .5811 | .4513 | .8245 |
| <u>Reading Comprehension</u> | | | | |
| F Value | 70.371 | 100.746 | 66.818 | 46.807 |
| probability>F | .0001 | .0001 | .0001 | .0001 |
| R-Squared (SAT-V explained variance) | .4059 | .4329 | .5269 | .2453 |
| Adj. R-Squared | .4001 | .4286 | .5190 | .2401 |
| Residuals (unexplained variance) | .5999 | .5714 | .4810 | .7599 |
| <u>Antonyms</u> | | | | |
| F Value | 103.530 | 98.440 | 72.599 | 88.236 |
| probability>F | .0001 | .0001 | .0001 | .0001 |
| R-Squared (SAT-V explained variance) | .5013 | .4272 | .5475 | .3799 |
| Adj. R-Squared | .4964 | .4228 | .5400 | .3756 |
| Residuals (unexplained variance) | .5036 | .5772 | .4600 | .6244 |
| <u>Quantitative Comparison</u> | | | | |
| F Value | 105.711 | 103.606 | 44.174 | 210.216 |
| probability>F | .0001 | .0001 | .0001 | .0001 |
| R-Squared (SAT-M explained variance) | .5065 | .4397 | .4240 | .5935 |
| Adj. R-Squared | .5017 | .4355 | .4144 | .5906 |
| Residuals (unexplained variance) | .4983 | .5645 | .5856 | .4094 |
| <u>Regular Math</u> | | | | |
| F Value | 77.507 | 136.436 | 28.126 | 133.840 |
| probability>F | .0001 | .0001 | .0001 | .0001 |
| R-Squared (SAT-M explained variance) | .4294 | .5083 | .3192 | .4817 |
| Adj. R-Squared | .4238 | .5045 | .3078 | .4781 |
| Residuals (unexplained variance) | .5762 | .4955 | .6922 | .5219 |
| <u>Data Interpretation</u> | | | | |
| F Value | 68.545 | 52.473 | 6.100 | 35.875 |
| probability>F | .0001 | .0001 | .0164 | .0001 |
| R-Squared (SAT-M explained variance) | .3996 | .2844 | .0923 | .1994 |
| Adj. R-Squared | .3937 | .2790 | .0772 | .1939 |
| Residuals (unexplained variance) | .6063 | .7210 | .9228 | .8061 |

| | | | | |
|---|---|---|---|---|
| **Analytic Reasoning** | | | | |
| F Value | 58.287 | 78.858 | 40.593 | 55.821 |
| probability>F | .0001 | .0001 | .0001 | .0001 |
| R-Squared (SAT-T explained variance) | .3614 | .3740 | .4035 | .2794 |
| Adj. R-Squared | .3552 | .3692 | .3936 | .2744 |
| Residuals (unexplained variance) | .6448 | .6308 | .6064 | .7256 |
| **Logical Reasoning** | | | | |
| F Value | 49.969 | 20.391 | 29.926 | 59.917 |
| probability>F | .0001 | .0001 | .0001 | .0001 |
| R-Squared (SAT-T explained variance) | .3267 | .1338 | .3328 | .3328 |
| Adj. R-Squared | .3201 | .1272 | .3217 | .2938 |
| Residuals (unexplained variance) | .6799 | .8728 | .6783 | .7062 |

## Regression analysis of GRE item-type scores on SAT sub-scores

The SAT scores explained smaller portions of variance in GRE item-type scores than in the GRE sub-scores (Verbal, Quantitative and Analytical). The SAT Verbal explained from 30 to 37 percent of the variance in the Sentence Completion item-type among the four college samples. The SAT Verbal explained from 18 to 55 percent of variation in the Analogies item-type, from 24 to 52 percent of the variation in Reading Comprehension and from 38 to 54 percent in Antonyms items. The SAT Math scores explained from 41 to 59 percent of variation in Quantitative Comparison item responses, from 31 to 50 percent of variation in Regular Math item-type scores, and from 8 to 39 percent of variation in Data Interpretation for the four samples. The regression of Data Interpretation scores on SAT Mathematics scores was the only procedure yielding a level of significance less than .0001 ($p$>.0164). The combined SAT Verbal and SAT Math scores (referred to as SAT Total) explained from 27 to 39 percent of variance in Analytic Reasoning and from 13 to 32 percent of variance in Logical Reasoning for the four samples. In all instances other than Data Interpretation, 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.

College 1 students entered college with slightly higher mean SAT Math score (699) than SAT Verbal score (638). As they approached graduation, these College 1 students remained <u>normatively</u> stronger in Quantitative abilities (mean GRE-Q = 701) than in Verbal abilities (mean GRE-V = 615). Yet, the regression analysis showed the College 1 students evinced large gains on specific GRE item-types that were not accounted for by the comparison of the means aggregates represented in the sub-scores of the two tests. Specific and significant gains were demonstrated in Analytic Reasoning a Logical Reasoning. These gains exceeded those found in the Sample #1 groups of the other institutions of the study. Such findings warrant further investigation of College 1 students general learned analytic abilities.

College 4 students entered college with slightly higher mean SAT Verbal scores (543) than SAT Math scores (523). As they approached graduation, these College 4 students remained <u>normatively</u> stronger in Verbal abilities (mean GRE-V = 522) than in Quantitative abilities (mean GRE-Q = 473). Yet, the regression analysis showed the College 4 students evinced large gains on specific GRE Quantitative item-types that were not accounted for by the comparison of the means aggregates represented in the test sub-scores. Specific and significant gains were demonstrated in Data Interpretation and Quantitative Comparisons. These gains exceeded those found in the college sample groups. Such findings warrant further investigation of College 4 students general learned in abilities in Data Interpretation and Quantitative Comparisons.

Using the student score gains obtained from the regression analysis above, the mean score gain for each course enrolling 5 or more students was calculated for all the 9 GRE item-types for each of the four colleges. Such a procedure did not assume that the specific gains of the students enrolled in each course were directly caused by that course. Rather, the score gains of each student were attributed to all the courses in which they enrolled, and the mean score gain for each course served as a proxy measure of student gains. Once courses are clustered by these gains, then hypotheses can be generated and tested as to why students who enrolled in a given pattern of courses experienced significant gains on one or more of the outcomes criteria (i.e., the item-type residuals). The procedures of the DCP Cluster Analytic Model, using these mean course score gains (residuals), is outlined in the following section.

## The DCP Project Cluster Analytic Model

### Cluster analytic model procedures

Described below are steps required in the Cluster analytic model 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' Scholastic Aptitude Test scores are used as variables to control for the academic abilities of these students when they first entered college. The students transcripts are used as the record of the sequence of courses 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 students transcripts which are associated with student score gains. This second objective is accomplished using cluster analysis, using their student score gains (GRE item-type residuals) as attributes of the courses in which they enrolled.

A raw data matrix consisting columns of courses and rows of score gains is created. The mean score gain for all the students in the sample who enrolled in a given course is calculated and used and becomes the metric value for that course. The correlation coefficient is used as the resemblance coefficient used to transform the data matrix into a resemblance matrix, wherein the similarity of score gains for students enrolling in one course can be compared with those enrolled in another course. Once the resemblance matrix indicating the proportional relationship of courses is established, then a clustering method is selected and executed to arra ge a tree or dendrogram of courses 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 courses have been correctly classified according to the 9 mean student score gains, b) to determine which of the 9 mean score gains were correlated with particular discriminant functions, and c) to determine which coursework clusters exhibited high mean score gains relative to each discriminant function. From the discriminant analysis, then, can be inferred an association between coursework patterns (clusters) and general learned abilities (student score gains on 9 criterion variables). The cluster-analytic procedure groups courses 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 analytic procedure:

Step 1. Calculate a student score gain 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 out. In this way, the student's academic abilities prior to entering college is controlled when calculating student score gains.

Step 2. Calculate the mean score gain for each course enrolling 5 or more students from the sample group. Cross-listed courses (courses with more than one possible course number) are standardized so that they have only one identifier. Courses with the same course identifier but with abstentiously different content (i.e., "Music 101: Voice" and "Music 101: Piano") are excluded from the analysis. Likewise 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 course for the purposes of analysis.

The proportion of courses included in the analysis is related to a) the extensiveness of the course 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 a sample groups will have enrolled in the same course. Likewise, the smaller the size of the student sample, the less frequently 5 or more students from a sample groups will have enrolled in the same course.

<u>Step 3</u>. Create a raw data matrix. The data matrix consists of columns representing of all the courses (objects) appearing on 5 or more student transcripts in the sample. The rows in the data matrix consist of the student gain (GRE item-type residual) scores. The mean residual score associated with that course and that student outcome measure is entered into each cell. For example, the course (object) in the first column in the data matrix is ANTHROPOLOGY 101, and 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 courses in the curriculum according to whether students who showed positive gains on each item-type were enrolled in the courses. This step prepares a raw data matrix to be used in a general cluster analysis based on quantitative data.

<u>Step 4</u>. 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 mat-

rix; it is expressed mathematically. There are many resemblance coefficients; each will express the similarity between courses (objects) in a slightly differently way. Each coefficient is appropriate for achieving slightly different research goals.

The resemblance coefficient selected is Pearson's product-moment correlation coefficient. It is appropriate for use with ratio data. For each course (object), it will calculate its similarity to each other course according to the 9 item-type residuals (attributes) coded in the data matrix. The resemblance coefficient expresses the relationship of two courses 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 correlatior resemblance coefficient. In this 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 course (object) in a pair, the rows represent the second course (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 course explain the variance in attributes on the second course. The resemblance coefficient serves a measure of similarity between one course and each other course 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 courses (objects) by use of a clustering method. A clustering method is a series of steps that removes values from the resemblance matrix, thereby reducing the size of the matrix. 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 inserted.

Romesburg (1984, p. 139) recommends using the unweighted pair-group method using arithmetic averages, also known as the average linkage method, which is abbreviated as UPGMA. 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 between courses 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 clusters and then conducting discriminant analyses on the resultant groupings, one can identify the number of clusters with has the highest predictive value, given the criterion variables used. Using the DISCRIMINANT 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 percentage of correct classification.

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 4-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 13-cluster solution may prove to be a 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.

<u>Step 8</u>. 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 function. Large positive and negative correlations are identified. 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.

Step 9. Repeat Steps 1 to 8 using a second cohort of students. Following Steps 1 through 8 will produced a set of hypothesized 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 courses according to their associated effect on student score gains. The Cluster Analytic Model classifies courses according to a ratio index of similarity to other courses. While the Cluster Analytic Model may examine only a fraction of all the courses in a college curriculum, it does provide a means to differentiate the effect of required courses or courses in which most student enroll. For example, a sample of 105 College 1 transcripts yielded 303 individual courses enrolling 5 or more of the students from a list of 1,445 unduplicated courses. While these 303 courses represented about one-fifth of all courses on the transcripts, a majority of the courses were those listed in the College 1 catalog as meeting the general education requirements of the University.

Creating the raw data matrix and the resemblance matrix

For each college sample, the mean score gains for each course were calculated. In College 3, for example, 149 courses were found on 5 or more of the student transcripts. A raw data matrix was created consisting of 149 columns and 9 rows (149 x 9). The rows represent the criterion variables: the mean student score gains on the 9 GRE item-type scores, while the columns represent those courses enrolling 5 or more students. Thus, each cell value of the matrix is a mean GRE item-type score gain for those sample group students enrolling in a specific course. The procedure was employed for each college sample.

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

For College 3, the resemblance matrix produced in this step consisted of 149 rows and 149 columns (149 x 149), in which each cell value theoretically ranges from -1.00 to 1.00. The calculation of the resemblance matrix was done using the SPSSx PROXIMITY program. The program provides 37 different proximity measures. Ten 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 numb-- 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.

Selection of the clustering method

The method selected for this quantitative analysis was the average linkage method (UPGMA). A summary of the results of the cluster analysis of the College 3 sample are presented in Figure 7. Courses were classified into 13 coursework patterns according to a hierarchical cluster structure. In fact, the choice to present the data in 13 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.

Using a 13 cluster solution to the quantitative cluster analysis, the largest number of courses are found in Coursework Cluster #1 with 44 courses. The smallest clusters are the 13th cluster with 1 course and the 7th, 10th and 11th clusters with 2 courses. 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 score gains on GRE item-types are strong, reliable and robust measures in differentiating student general learned abilities.

In the cluster analysis developed for each of the four college samples, a careful examination of courses within each cluster indicated that some courses from the same department tend to be combined together. Also, a set of courses coming from various related disciplines may form a homogeneous cluster on the basis of a set of given attributes or criteria. At this point in the analysis, it was difficult to describe which dimensions of student general learned ability each clusters represents. However, it seemed to be clear that one pattern of course enrollment may contribute to student general learned ability in a way significantly different from the other coursework patterns and that those patterns varied from college to college. Supporting this is a more detailed examination of subset courses of each clusters. In many cases, those courses offered at the same level tend to be combined into pairs together. But, those pairs are agglomerated with other courses offered at the higher level again according to the hierarchical structure of clusters. This may suggest that student gains in general learned abilities may be obtained through a sequential enrollment pattern during the college years, not at a single stage of the sequence (such as the freshman year experience).

A sample hierarchical cluster structure is presented for College 3 in the dendrogram of Figure 7b. For concise visual presentation, the complex sub-structures of each of the clusters were omitted from the dendrogram in Figure 7b. The dendrogram displays the clusters being combined and the distances between the clusters at each successive steps, suggesting that the 13-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 between courses), 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.

Figure 7-b. Dendrogram of hierarchical cluster structure for College 3

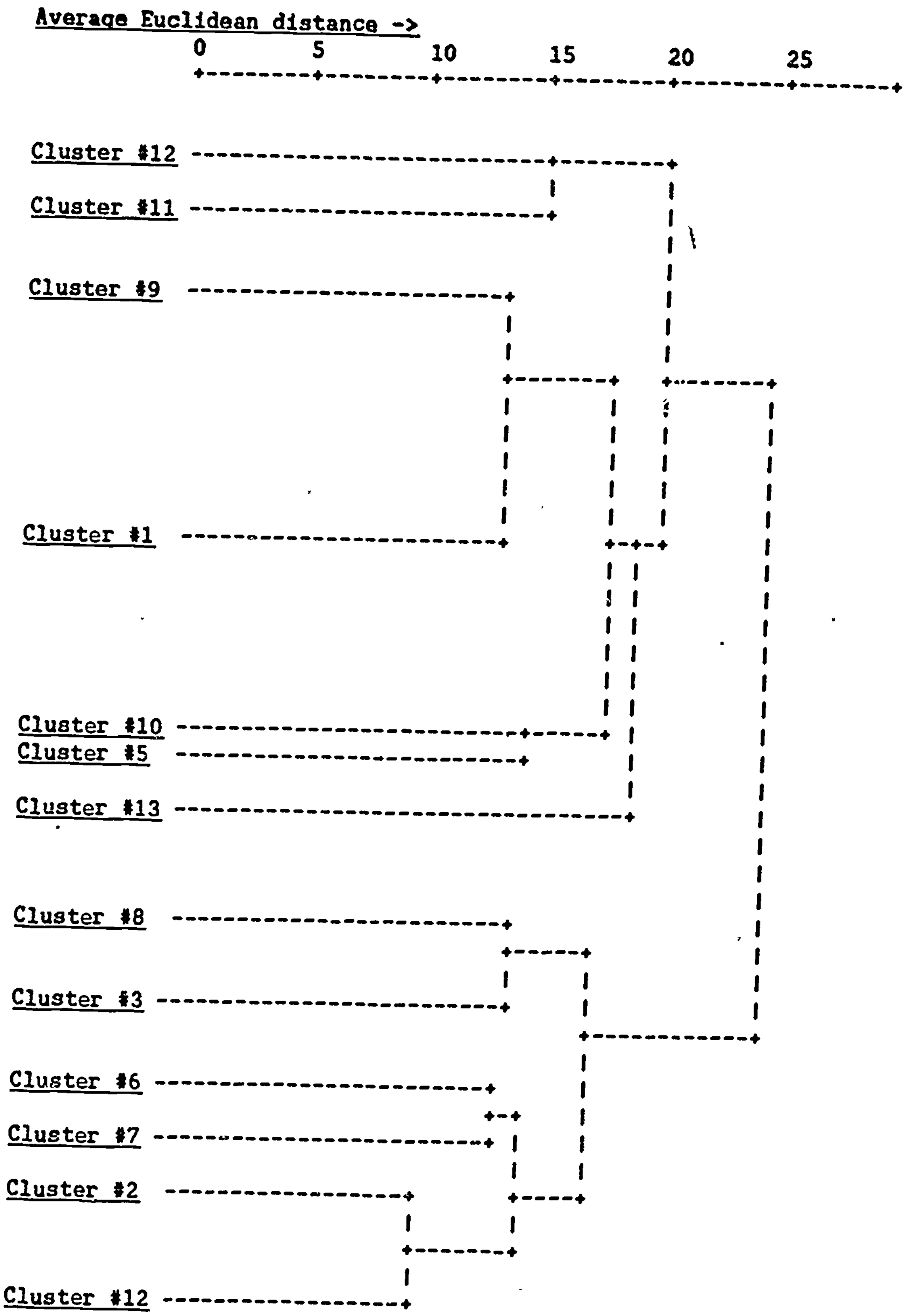


Figure 7-c. Dendrogram of hierarchical cluster structure for College 3.

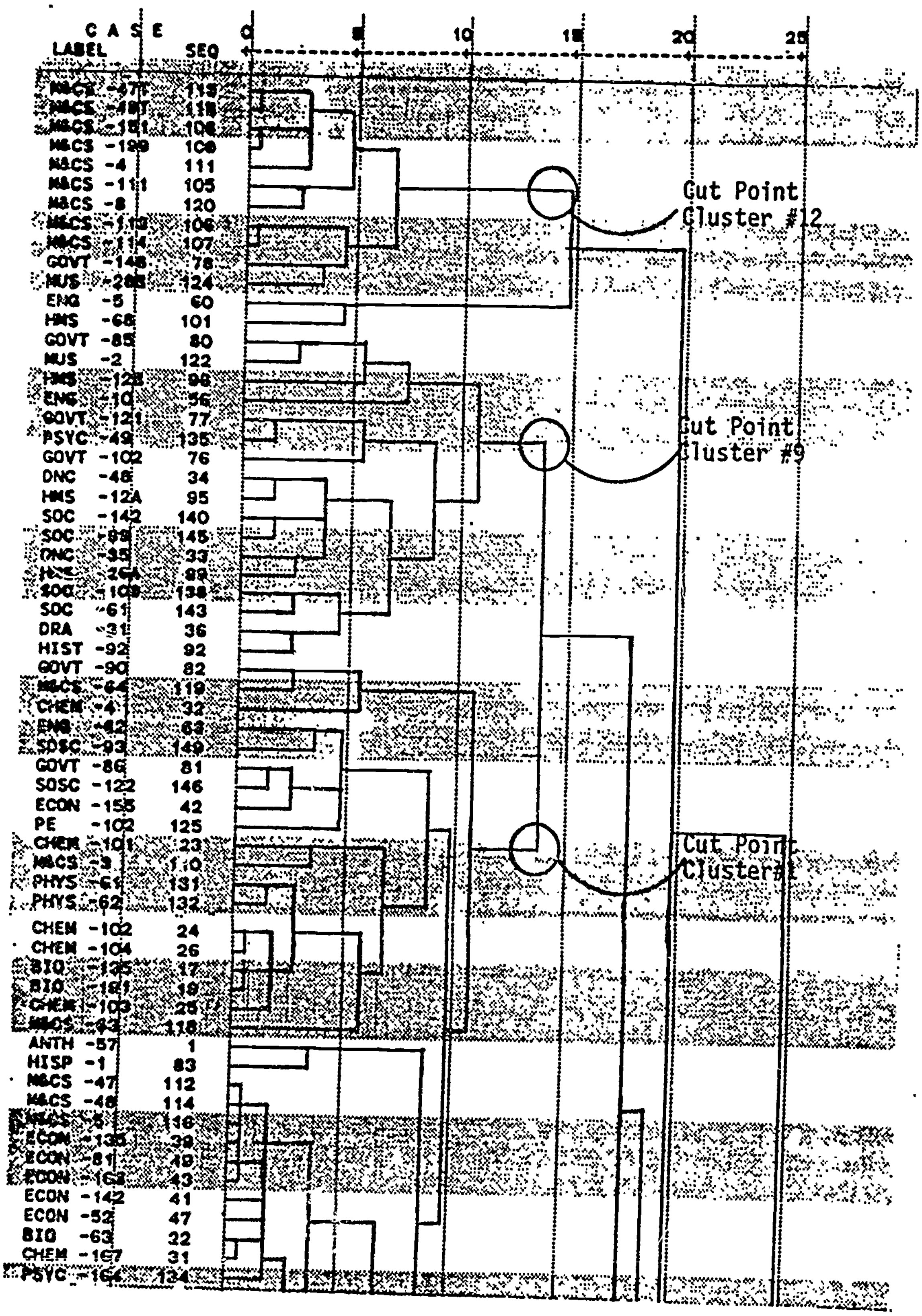


Figure 7-c. Dendrogram of hierarchical cluster structure for College 3.

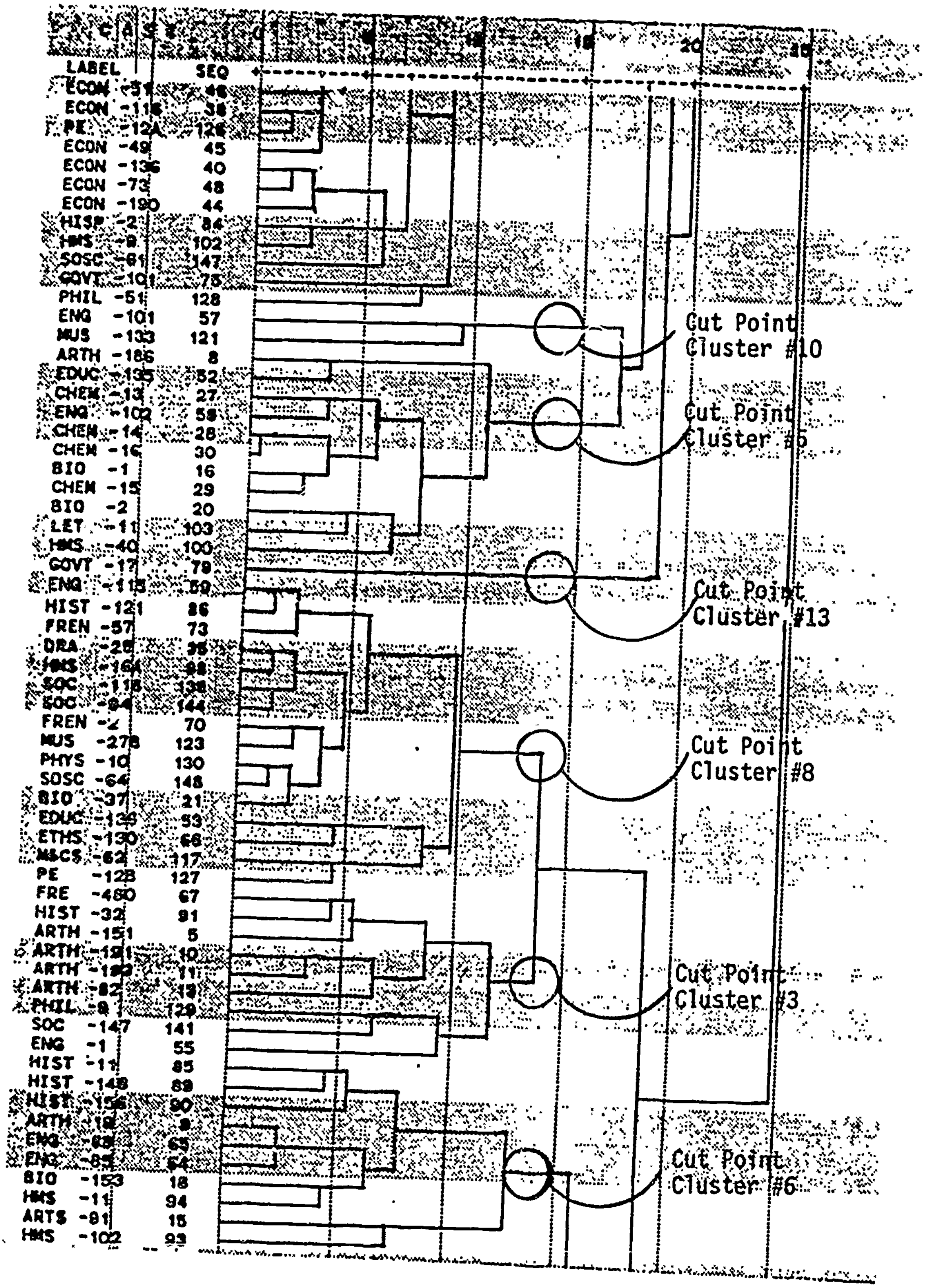


Figure 7-c. Dendrogram of Hierarchical cluster structure for College 3.

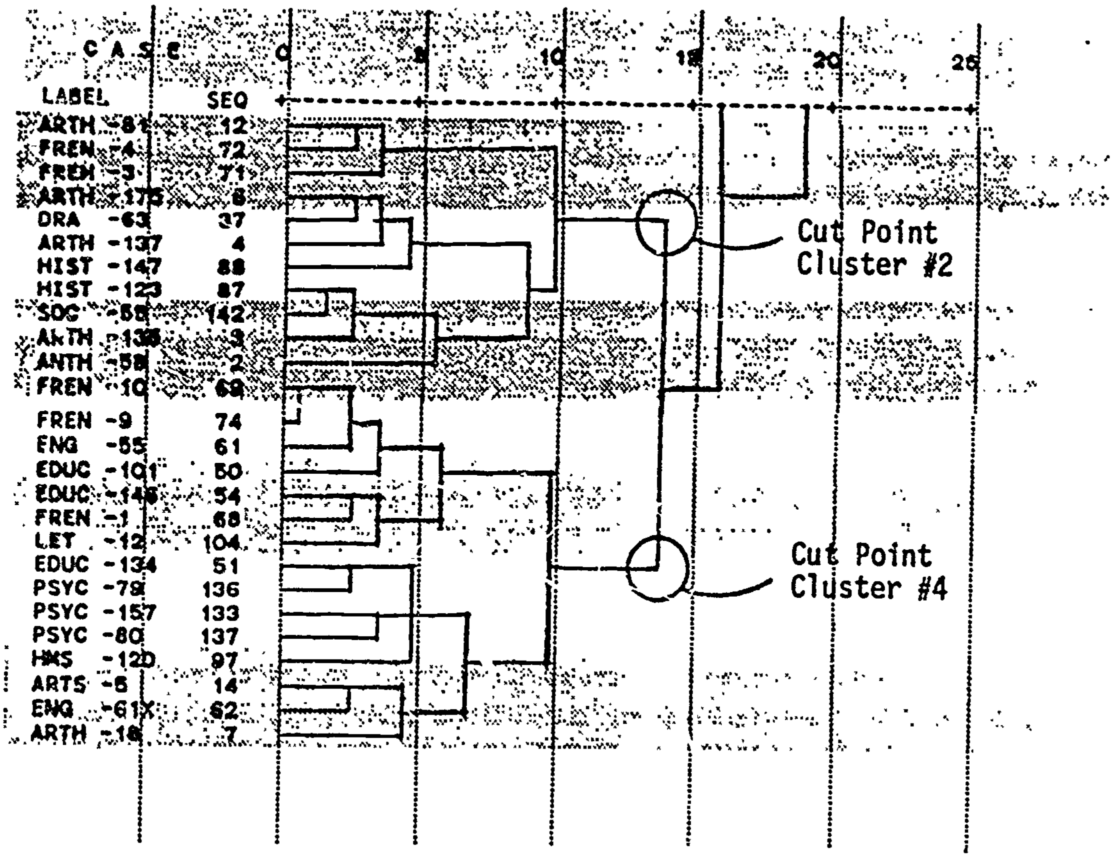


Figure 8a. Courses within coursework clusters (13-cluster solution): College 3

| Cluster #1 n = 44 | Cluster #2 n = 11 | Cluster #3 n = 9 | Cluster #4 n = 15 | Cluster #5 n = 11 |
|---|---|---|---|---|
| ANTH 57 | ANTH 58 | ARTH 151 | ARTH 18 ** | ARTH 126 |
| BIO 135 | ARTH 135 | ARTH 191 | ARTS 5 ** | BIO 1 |
| BIO 191 | ARTH 137 | ARTH 193 | EDUC 101 | BIO 2 |
| BIO 63 | ARTH 175 | ARTH 82 | EDUC 134 | CHEM 13 |
| CHEM 101 | ARTH 81 | ENG 1 | EDUC 146 | CHEM 14 |
| CHEM 102 | DRA 63 ** | FRE 480 | ENG 55 | CHEM 15 |
| CHEM 103 | FREN 3 | HIST 32 | ENG 61X | CHEM 16 |
| CHEM 104 | FREN 4 | PHIL 9 | FREN 1 | EDUC 135 |
| CHEM 167 | HIST 123 | SOC 147 | FREN 10 | ENG 102 |
| CHEM 4 ** | HIST 147 | | FREN 9 | HMS 40 |
| ECON 116 | SOC 55 | | HMS 12D | LET 11 |
| ECON 135 | | | LET 12 | |
| ECON 136 | | | PSYC 157 | |
| ECON 142 | | | PSYC 79 | |
| ECON 155 | | | PSYC 80 | |
| ECON 163 | | | | |
| ECON 190 | | | | |
| ECON 49 | | | | |
| ECON 51 | | | | |
| ECON 52 | | | | |
| ECON 73 | | | | |
| ECON 81 | | | | |
| ENG 62 | | | | |
| GOVT 101 | | | | |
| GOVT 86 | | | | |
| GOVT 90 | | | | |
| HISP 1 | | | | |
| HISP 2 | | | | |
| HMS 9 | | | | |
| M&CS 3 | | | | |
| M&CS 47 | | | | |
| M&CS 48 | | | | |
| M&CS 5 | | | | |
| M&CS 63 | | | | |
| M&CS 64 | | | | |
| PE 102 | | | | |
| PE 12A | | | | |
| PHIL 51 | | | | |
| PHYS 61 | | | | |
| PHYS 62 | | | | |
| PSYC 164 | | | | |
| SOSC 122 | | | | |
| SOSC 61 | | | | |
| SOSC 93 | | | | |

"*" following a course indicates a course misclassified according to the discriminant analysis of course clusters.

Figure 8b. Courses within coursework clusters (13-cluster solution): College 3

| Cluster #6 n = 8 | Cluster #8 n = 16 | Cluster #9 n = 17 | Cluster #10 n = 2 | Cluster #12 n = 11 |
|---|---|---|---|---|
| ARTH 19 | BIO 37 | DNC 35 | ENG 101 ** | GOVT 148 |
| BIO 153 | DRA 25 | DNC 48 | MUS 133 | M&CS 111 |
| ENG 85 | EDUC 136 | DRA 31 | | M&CS 113 |
| ENG 88 | ENG 115 | ENG 10 | | M&CS 114 |
| HIST 11 | ETHS 130 | GOVT 102 | **Cluster #11** | M&CS 151 |
| HIST 148 | FREN 2 | GOVT 148 | n = 2 | M&CS 199 |
| HIST 156 | FREN 57 | GOVT 85 | | M&CS 4 |
| HMS 11 | HIST 121 | HIST 92 ** | ENG 5 | M&CS 47T |
| | HMS 16A | HMS 12A | HMS 68 | M&CS 48T |
| | M&CS 62 | HMS 12B | | M&CS 8 |
| **Cluster #7** | MUS 27B | HMS 26A | | MUS 28B |
| n = 2 | PE 12B** | MUS 2 ** | | |
| | PHYS 10 | PSYC 49 | | |
| ARTS 91 | SOC 116 | SOC 103 | | |
| HMS 102 | SOC 94 | SOC 142 | | **Cluster #13** |
| | SOSC 64 | SOC 61 | | n = 1 |
| | | SOC 99 | | |
| | | | | GOVT 17 |

"*" following a course indicates a course misclassified according to the discriminant analysis of course clusters.

A careful examination of courses within each cluster seems to indicate that some courses coming from the same department may appear in the same cluster, such as the Economics courses (Econ) in Cluster #1. Similarly, there are apparent sequences of courses, such as the Chemistry 101, 102, 103, 104 sequence in Cluster #1. Also, a set of courses coming from various related disciplines may form a homogeneous cluster on the basis of a set of given attributes or criteria. There appears to be a homogeneity of humanities and social science disciplines in Cluster #2.

At this point in the analysis, it is difficult to describe which dimensions of student general learned ability are represented in each cluster. However, it seems to be clear that one pattern of course enrollment may contribute to student general learned ability in a way significantly different from the other coursework patterns. Supporting this is a more detailed examination of subset courses of each clusters. In many cases, those courses offered at the same level tend to be combined into pairs together. But, those pairs are agglomerated with other courses offered at the higher level again according to the hierarchical structure of clusters. This may suggest that student gains in general learned abilities may be obtained through a sequential enrollment pattern during the college years, not at a single stage of the sequence (such as the freshman year experience). These observations were repeatedly conveyed in the analysis of each of the four college samples.

Discriminant analysis of coursework patterns

In examining the dendrogram of coursework, a logical question arises as to which number of clusters or pattern groupings provides the best explanation of the relationship between student item-type gain scores and coursework patterns. Separate discriminant analyses of different numbers of cluster groupings were performed for each college sample in order to determine the number of groupings that optimizes the proportion of courses correctly classified. Four different cluster solutions provided comparably high levels of correct classification (See Figure 9):

Figure 9. Percent of correct classification by college and cluster solution.

| | Percent of courses correctly classified according to Discriminant Analysis of Cluster solution | | | |
|---|---|---|---|---|
| Cluster solution | College 1 | College 2 | College 3 | College 4 |
| 5 cluster solution : | | 89.92% | 89.93% | 87.65% |
| 8 cluster solution : | 82.48% | 90.07% | 92.99% | 87.41% |
| 13 cluster solution : | 85.48% | 89.36% | 94.63% | 85.43% |
| 25 cluster solution : | 83.83% | 86.17% | 90.45% | 83.70% |

Accuracy versus explanation

While these cluster solutions produced comparable classification results, the different grouping evidenced differing effectiveness in identifying relationships between mean item-type residuals (gain scores) and coursework patterns. The 13-cluster solution proved to provide the greatest extent of information about such relationships and was therefore used in this report.

The discriminant analysis was conducted using the DISCRIMINANT program in SPSSx in the following manner. Using the course 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 courses 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). As Figure 9 demonstrates, the DCP Cluster Analytic Model consistently grouped courses correctly according to the 9 criterion variables. Figure 10 provides an example of the classification breakdown for College 3 as an example.

Figure 10. Discriminant analysis of the 13-cluster solution for College 3.

| Actual Cluster | No. of Cases | Predicted Group Membership Gr 1 | Gr 2 | Gr 3 | Gr 4 | Gr 5 | Gr 6 | Gr 7 | Gr 8 | Gr 9 | Gr 10 | Gr 11 | Gr 12 | Gr 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group 1 | 44 | 43 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| | | 97.7% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | 2.3% | .0% | .0% | .0% | .0% |
| Group 2 | 11 | 0 | 10 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | .0% | 90.9% | .0% | 9.1% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% |
| Group 3 | 9 | 0 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | .0% | .0% | 100.0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% |
| Group 4 | 15 | 0 | 2 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | .0% | 13.3% | .0% | 86.7% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% |
| Group 5 | 11 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | .0% | .0% | .0% | .0% | 100.0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% |
| Group 6 | 8 | 0 | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | .0% | .0% | .0% | .0% | .0% | 100.0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% |
| Group 7 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | .0% | .0% | .0% | .0% | .0% | .0% | 100.0% | .0% | .0% | .0% | .0% | .0% | .0% |
| Group 8 | 16 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 |
| | | .0% | .0% | .0% | .0% | 6.3% | .0% | .0% | 93.8% | .0% | .0% | .0% | .0% | .0% |
| Group 9 | 17 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 15 | 0 | 0 | 0 | 0 |
| | | 5.9% | .0% | .0% | .0% | .0% | .0% | .0% | 5.9% | 88.2% | .0% | .0% | .0% | .0% |
| Group 10 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | .0% | .0% | .0% | .0% | 50.0% | .0% | .0% | .0% | .0% | 50.0% | .0% | .0% | .0% |
| Group 11 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| | | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | 100.0% | .0% | .0% |
| Group 12 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 |
| | | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | 100.0% | .0% |
| Group 13 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | .0% | 100.0% |

Percent of "Grouped" Clusters correctly classified: 94.63%

Correlations of item-types and discriminant functions

The discriminant analysis of College 3 Sample #1 provided secondary validation that 94.63% of the classification of courses was correctly predicted by the cluster analysis (Figure 10). The discriminant analysis is a secondary validation, since it is based on the same sample of transcripts and test scores; primary validation of the cluster analysis may be achieved through the replication of coursework patterns in the College 3 Sample #2 to be analyzed in the fall of 1989.

Stated simply, over 9 of 10 courses more frequently taken by College 3 Sample #1 students were correctly classified according to their mean residual GRE scores. While the cluster analysis produces coursework patterns according to criteria of general student learning, additional steps are needed (1) to determine which courses were correctly classified and (2) to ascertain which item-type scores contributed to any given coursework pattern.

Using the BREAKDOWN procedure in the DISCRIMINANT program of SPSS-X (Norusis, 1985), courses which were incorrectly classified or which may be classified within another coursework pattern are identified. When a set of data is taxonomized by multiple criteria, it is reasonable to expect that the first groupings are generally the most homogeneous and the later groupings are the most heterogeneous. Consequently, for College 3, the highest proportions of misclassifications occurred in Cluster #10, while all other Clusters had relatively few misclassifications (See Figures 8a-8b, 9).

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.

Correlations of coursework clusters and discriminant functions

Examination of the relationship between GRE item-type residuals and discriminant functions revealed the following pooled within-group correlations over -/+ .50:

Function 1 was positively correlated to Quantitative Comparisons ($r$=.61);
Function 2 was positively correlated to Analytic Reasoning ($r$=.77);
Function 3 was positively correlated to Antonyms ($r$=.64);
Function 4 evinced no strong correlations;
Function 5 was positively correlated to Data Interpretation ($r$=.52), but was negatively correlated to Sentence Completion ($r$=-.64);
Function 6 evinced no strong correlations;
Function 7 was positively correlated to Analogies ($r$=.52);
Function 8 was positively correlated to Regular Mathematics ($r$=.72);
Function 9 was positively correlated to Quantitative Comparisons ($r$=.54), was positively correlated to Reading Comprehension ($r$=.63), and was positively correlated to Logical Reasoning ($r$=.75).

Once the relationships between discriminant functions and mean item-type residuals have been established, then the relationships between the discriminant functions and the coursework clusters can also be determined.

By examining the average score of each cluster group for each discriminant function, the extent to which each discriminant function contributes to that group may be calculated. The average residual score for a coursework cluster group is called the group centroid. Group centroids for each coursework cluster in the College 3 are presented in Figure 11 as an example.

Figure 11. Canonical discriminant functions evaluated at group means: College 3.

| Cluster | Func 1 | Func 2 | Func 3 | Func 4 | Func 5 | Func 6 | Func 7 | Func 8 | Func 9 |
|---|---|---|---|---|---|---|---|---|---|
| Cluster #1 | -1.8550 | 1.1405 | .8196 | .4000 | .4852 | -.1604 | -.4737 | -.1180 | .0984 |
| Cluster #2 | 1.9477 | -.2831 | -1.0923 | -1.0480 | -.0623 | .7887 | -.4906 | .0360 | -.0335 |
| Cluster #3 | 1.7908 | .0112 | .6087 | -.8789 | -.0236 | -.6195 | -1.2068 | -.0313 | .3472 |
| Cluster #4 | 1.0542 | -2.0374 | -1.7323 | -.3420 | -.8559 | 1.0323 | 1.3298 | .5855 | .1410 |
| Cluster #5 | -1.4100 | .3401 | -1.1924 | .8094 | -.5168 | -.5452 | .5356 | .3206 | -.4135 |
| Cluster #6 | -.3076 | .4274 | -1.5957 | -1.4521 | .4679 | .7115 | .3007 | 2.1162 | -.0851 |
| Cluster #7 | .0242 | .7631 | -2.4056 | .3669 | -1.6100 | .5107 | -.7840 | -.1865 | 1.2316 |
| Cluster #8 | 2.3077 | -2.3788 | .1281 | .7917 | -1.1630 | .8293 | .0686 | .4798 | -.2749 |
| Cluster #9 | -.0228 | -.8669 | 1.8940 | .4101 | -.0506 | -.6678 | .3541 | -.1248 | -.1877 |
| Cluster #10 | .3681 | -.8543 | -1.2248 | 2.3499 | 2.2352 | -2.2273 | -.9445 | -1.0648 | 2.7481 |
| Cluster #11 | -.3361 | 1.2296 | 1.2284 | -2.5238 | 1.7931 | .1739 | -1.1577 | .3300 | 1.8070 |
| Cluster #12 | .9251 | 2.5221 | -.6097 | -.8765 | .5369 | -.8357 | .6265 | -.5458 | -.5968 |
| Cluster #13 | -.5869 | -.9901 | -.3927 | -.7007 | 1.3323 | -.8766 | .6203 | .5365 | -1.3882 |
| Minimum | -1.8550 | -2.3788 | -2.4056 | -2.5238 | -1.6100 | -2.2273 | -1.2068 | -1.0648 | -1.3882 |
| Maximum | 2.3077 | 2.5221 | 1.8940 | 2.3499 | 2.2352 | 1.0323 | 1.3298 | 2.1162 | 2.7481 |
| Average | .2999 | -.0751 | -.4282 | -.2072 | .1976 | -.1436 | -.0940 | .1795 | .2610 |

Interpreting the coursework clusters for the 13-cluster solution

Figure 11 shows the coursework cluster means (group centroids) for each discriminant function for College 3. Clusters with positive or negative means greater than 1.0 were selected for further analysis.

Coursework Cluster #1 had a positive group mean on Function 2 and a high negative group mean on Function 1. Function 1 was positively correlated to Quantitative Comparisons ($r$=.61), and Function 2 was positively correlated to Analytic Reasoning ($r$=.77).Students enrolling in the coursework of Cluster #1 showed improvement in Quantitative Comparisons and decline in Analytic Reasoning abilities.

In contrast, Cluster #2 had a high positive group mean on Function 1. This Cluster also evinced high negative group means on Function 3 and Function 4. Function 3 was positively correlated to Antonyms ($\underline{r}$=.64), while Function 4 exhibited no strong correlations. Students enrolling in Cluster #2 coursework showed improvement in their Quantitative Comparisons abilities, but dipped in their ability to answer Antonyms questions correctly.

Cluster #3 had a high positive group mean on Function 1 and a high negative group mean on Function 7. Function 1 was positively correlated to Quantitative Comparisons ($\underline{r}$=.61). Function 7 was positively correlated to Analogies ($\underline{r}$=.52). Students enrolled in this coursework cluster showed improvement in Quantitative Comparisons but declined in Analogies.

Cluster #4 had a high positive group means on Functions 1, 6, and 7. Cluster #4 also evidenced high negative group means on Functions 2 and 3. Function 1 was positively correlated to Quantitative Comparisons ($\underline{r}$=.61). Function 2 was positively correlated to Analytic Reasoning ($\underline{r}$=.77), while Function 3 was positively correlated to Antonyms ($\underline{r}$=.64). Function 6 evinced no strong correlations. Function 7 was positively correlated to Analogies ($\underline{r}$=.52). Therefore, students enrolling in this coursework bettered their performance in Quantitative Comparisons and Analogies, but flagged in Antonyms and Analytic Reasoning.

Cluster #5 had a high negative group mean on Functions 1 and 3. Accordingly, students who enrolled in the coursework pattern represented in Cluster #5 displayed declines in their ability to make Quantitative Comparisons and to decipher Analogies.

Cluster #6 had a high positive group mean on Function 8 and high negative group means on Function 3 and 4. Function 3 was positively correlated to

Antonyms ($r$=.64), while Function 4 exhibited no strong correlations. Function 8 was positively correlated to Regular Mathematics ($r$=.72). Students enrolled in Cluster #6 coursework showed gains in Regular Mathematics but evinced decline in their ability to identify Antonyms.

Cluster #7 had a high positive group mean on Function 9 and high negative group means on Functions 3 and 5. Function 3 was positively correlated to Antonyms ($r$=.64). Function 5 was positively correlated to Data Interpretation ($r$=.52), but was negatively correlated to Sentence Completion ($r$=-.64). Function 9 was positively correlated to Quantitative Comparisons ($r$=.54), Reading Comprehension ($r$=.63), and Logical Reasoning ($r$=.75). Consequently, students registered in Cluster #7 coursework demonstrated gains in Reading Comprehension, Sentence Completion, Quantitative Comparisons and Logical Reasoning. They also showed declines in abilities in Antonyms and Data Interpretation. It should be noted that Cluster #7 had but 2 members, albeit that the discriminant analysis indicated that both were correctly classified (See Figure 9).

Cluster #8 produced a positive high group mean on Function 1 and high negative group means on Functions 2 and 5. Function 1 was positively correlated to Quantitative Comparisons ($r$=.61), while Function 2 was positively correlated to Analytic Reasoning ($r$=.77). Function 5 was positively correlated to Data Interpretation ($r$=.52), but was negatively correlated to Sentence Completion ($r$=-.64). Students listed on the rolls of Cluster #8 coursework evinced improvement in Data Interpretation Quantitative Comparisons and Sentence Completion. They exhibited declines in their ability to answer Data Interpretation and Analytic Reasoning questions.

Cluster #9 showed a high positive group mean on Function 3. Function 3 was positively correlated to Antonyms ($r$=.64). Students in this coursework improved in their abilities on this GRE item-type.

**Cluster #10** had only 2 courses as members, one of which was misclassified (See Figure 9). Under such circumstances further analysis was not appropriate.

**Cluster #11** offered a high positive group means on Function 2, 3, 5, 7 and 9 and a high negative group mean on Function 4. While Cluster #11 had only 2 members, both were correctly classified (See Figure 9). Function 2 was positively correlated to Analytic Reasoning (*r*=.77), while Function 3 was positively correlated to Antonyms (*r*=.64). Function 4 exhibited no strong correlations. Function 5 was positively correlated to Data Interpretation (*r*=.52), but was negatively correlated to Sentence Completion (*r*=-.64). Function 7 was positively correlated to Analogies (*r*=.52). Function 9 was positively correlated to Quantitative Comparisons (*r*=.54), Reading Comprehension (*r*=.63), and Logical Reasoning (*r*=.75). Students enrolling in Cluster #11 coursework improved in their ability to correctly answer Antonyms, Analogies, Reading Comprehension, Quantitative Comparisons, Data Interpretation, Analytic Reasoning, and Logical Reasoning questions, but dipped in their Sentence Completion abilities.

**Cluster #12** had a high positive group mean on Function 2. Function 2 was positively correlated to Analytic Reasoning (*r*=.77). Students enrolled in this Cluster improved in their ability in correctly answering Analytic Reasoning questions.

**Cluster #13** had but one member, albeit that member was correctly classified. Cluster 13 had a high negative group mean on Function 9 and a high positive group mean on Function 5. Function 5 was positively correlated to Data Interpretation (*r*=.52), but was negatively correlated to Sentence Completion (*r*=-.64). Function 7 was positively correlated to Analogies (*r*=.52). Function 9 was positively correlated to Quantitative Comparisons (*r*=.54), Reading Comprehension (*r*=.63), and Logical Reasoning (*r*=.75). Students enrolled in this coursework cluster declined in Reading Comprehension, Quantitative Comparisons and Logical Reasoning and improved in Data Interpretation.

Figures 12a - 12f portray the College 3 coursework clusters arranged according to the GRE item-type with which they were associated. It should be cautioned that the association was established at the <u>cluster</u> level. No direct causal link was intimated between student enrollment in any one given course and scores on the GRE. Furthermore, at this point, one cannot say why students who enrolled in these courses had higher score gains. The cluster serves to hypothesize relationships between coursework 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 coursework tended to perform better on specific item-types within the GRE, while others who enrolled in different coursework patterns did not tend to perform as well. Comparable analyses of the other college samples produced similar relationships.

Figure 12a. Coursework clusters associated with high positive mean residuals.

**High Pos. Mean Residuals on Analogies**

Cluster #4
n = 15

| | |
|---|---|
| ARTH | 18 * |
| ARTS | 5 * |
| EDUC | 101 |
| EDUC | 134 |
| EDUC | 146 |
| ENG | 55 |
| ENG | 61X |
| FREN | 1 |
| FREN | 10 |
| FREN | 9 |
| HMS | 12D |
| LET | 12 |
| PSYC | 157 |
| PSYC | 79 |
| PSYC | 80 |

Cluster #11
n = 2

| | |
|---|---|
| ENG | 5 |
| HMS | 68 |

**High Pos. Mean Residuals on Sentence Completion**

Cluster #7
n = 2

| | |
|---|---|
| ARTS | 91 |
| HMS | 102 |

Cluster #8
n = 16

| | |
|---|---|
| BIO | 37 |
| DRA | 25 |
| EDUC | 136 |
| ENG | 115 |
| ETHS | 130 |
| FREN | 2 |
| FREN | 57 |
| HIST | 121 |
| HMS | 16A |
| M&CS | 62 |
| MUS | 27B |
| PE | 12B * |
| PHYS | 10 |
| SOC | 116 |
| SOC | 94 |
| SOSC | 64 |

**High Pos. Mean Residuals on Reading Comprehension**

Cluster #7
n = 2

| | |
|---|---|
| ARTS | 91 |
| HMS | 102 |

Cluster #10
n = 2

| | |
|---|---|
| ENG | 101 * |
| MUS | 133 |

Cluster #11
n = 2

| | |
|---|---|
| ENG | 5 |
| HMS | 68 |

**High Pos. Mean Residuals on Antonyms**

Cluster #9
n = 17

| | |
|---|---|
| DNC | 35 |
| DNC | 48 |
| DRA | 31 |
| ENG | 10 |
| GOVT | 102 |
| GOVT | 148 |
| GOVT | 85 |
| HIST | 92 * |
| HMS | 12A |
| HMS | 12B |
| HMS | 26A |
| MUS | 2 * |
| PSYC | 49 |
| SOC | 103 |
| SOC | 142 |
| SOC | 61 |
| SOC | 99 |

Cluster #11
n = 2

| | |
|---|---|
| ENG | 5 |
| HMS | 68 |

"*" following a course indicates a course misclassified according to the discriminant analysis of course clusters.

Figure 12b. Coursework clusters associated with high positive mean residuals.

## High Pos. Mean Residuals on Quantitative Comparisons

**Cluster #2**
n = 11

| | |
|---|---|
| ANTH | 58 |
| ARTH | 135 |
| ARTH | 137 |
| ARTH | 175 |
| ARTH | 81 |
| DRA | 63 * |
| FREN | 3 |
| FREN | 4 |
| HIST | 123 |
| HIST | 147 |
| SOC | 55 |

**Cluster #3**
n = 9

| | |
|---|---|
| ARTH | 151 |
| ARTH | 191 |
| ARTH | 193 |
| ARTH | 82 |
| ENG | 1 |
| FRE | 480 |
| HIST | 32 |
| PHIL | 9 |
| SOC | 147 |

**Cluster #4**
n = 15

| | |
|---|---|
| ARTH | 18 * |
| ARTS | 5 * |
| EDUC | 101 |
| EDUC | 134 |
| EDUC | 146 |
| ENG | 55 |
| ENG | 61X |
| FREN | 1 |
| FREN | 10 |
| FREN | 9 |
| HMS | 12D |
| LET | 12 |
| PSYC | 157 |
| PSYC | 79 |
| PSYC | 80 |

**Cluster #7**
n = 2

| | |
|---|---|
| ARTS | 91 |
| HMS | 102 |

**Cluster #8**
n = 16

| | |
|---|---|
| BIO | 37 |
| DRA | 25 |
| EDUC | 136 |
| ENG | 115 |
| ETHS | 130 |
| FREN | 2 |
| FREN | 57 |
| HIST | 121 |
| HMS | 16A |
| M&CS | 62 |
| MUS | 27B |
| PE | 12B* |
| PHYS | 10 |
| SOC | 116 |
| SOC | 94 |
| SOSC | 64 |

**Cluster #10**
n = 2

| | |
|---|---|
| ENG | 101 * |
| MUS | 133 |

**Cluster #11**
n = 2

| | |
|---|---|
| ENG | 5 |
| HMS | 68 |

## High Pos. Mean Residuals on Regular Mathematics

**Cluster #6**
n = 8

| | |
|---|---|
| ARTH | 19 |
| BIO | 153 |
| ENG | 85 |
| ENG | 88 |
| HIST | 11 |
| HIST | 148 |
| HIST | 156 |
| HMS | 11 |

## High Pos. Mean Residuals on Data Interpretation

**Cluster #10**
n = 2

| | |
|---|---|
| ENG | 101 * |
| MUS | 133 |

**Cluster #11**
n = 2

| | |
|---|---|
| ENG | 5 |
| HMS | 68 |

**Cluster #13**
n = 1

| | |
|---|---|
| GOVT | 17 |

"*" following a course indicates a course misclassified according to the discriminant analysis of course clusters.

Figure 12c. Coursework clusters associated with high positive mean residuals.

## High Pos. Mean Residuals on Analytic Reasoning

**Cluster #1**, n = 44

| | |
|---|---|
| ANTH | 57 |
| BIO | 135 |
| BIO | 191 |
| BIO | 63 |
| CHEM | 101 |
| CHEM | 102 |
| CHEM | 103 |
| CHEM | 104 |
| CHEM | 167 |
| CHEM | 4 * |
| ECON | 116 |
| ECON | 135 |
| ECON | 136 |
| ECON | 142 |
| ECON | 155 |
| ECON | 163 |
| ECON | 190 |
| ECON | 49 |
| ECON | 51 |
| ECON | 52 |
| ECON | 73 |
| ECON | 81 |
| ENG | 62 |
| GOVT | 101 |
| GOVT | 86 |
| GOVT | 90 |
| HISP | 1 |
| HISP | 2 |
| HMS | 9 |
| M&CS | 3 |
| M&CS | 47 |
| M&CS | 48 |
| M&CS | 5 |
| M&CS | 63 |
| M&CS | 64 |
| PE | 102 |
| PE | 12A |

**Cluster #1**, n = 44

| | |
|---|---|
| PHIL | 51 |
| PHYS | 61 |
| PHYS | 62 |
| PSYC | 164 |
| SOSC | 122 |
| SOSC | 61 |
| SOSC | 93 |

**Cluster #11**, n = 2

| | |
|---|---|
| ENG | 5 |
| HMS | 68 |

**Cluster #12**, n = 11

| | |
|---|---|
| GOVT | 148 |
| M&CS | 111 |
| M&CS | 113 |
| M&CS | 114 |
| M&CS | 151 |
| M&CS | 199 |
| M&CS | 4 |
| M&CS | 47T |
| M&CS | 48T |
| M&CS | 8 |
| MUS | 28B |

## High Pos. Mean Residuals on Logical Reasoning

**Cluster #7**, n = 2

| | |
|---|---|
| ARTS | 91 |
| HMS | 102 |

**Cluster #10**, n = 2

| | |
|---|---|
| ENG | 101 |
| MUS | 133 |

**Cluster #11**, n = 2

| | |
|---|---|
| ENG | 5 |
| HMS | 68 |

"*" following a course indicates a course misclassified according to the discriminant analysis of course clusters.

Figure 12d. Coursework clusters associated with high negative mean residuals.

| High Neg. Mean Residuals on Analogies | High Neg. Mean Residuals on Sentence Completion | High Neg. Mean Residuals on Reading Comprehension |
|---|---|---|
| <u>Cluster #3</u> | <u>Cluster #10</u> | <u>Cluster #13</u> |
| n = 9 | n = [illegible] | n = 1 |
| ARTH 151 | ENG 101 * | GOVT 17 |
| ARTH 191 | MUS 133 | |
| ARTH 193 | | |
| ARTH 82 | | |
| ENG 1 | <u>Cluster #11</u> | |
| FRE 480 | n = 2 | |
| HIST 32 | | |
| PHIL 9 | ENG 5 | |
| SOC 147 | HMS 58 | |

"*" following a course indicates a course misclassified according to the discriminant analysis of course clusters.

Figure 12e. Coursework clusters associated with high negative mean residuals.

## High Neg. Mean Residuals on Antonyms

**Cluster #2**
n = 11

| | |
|---|---|
| ANTH | 58 |
| ARTH | 135 |
| ARTH | 137 |
| ARTH | 175 |
| ARTH | 81 |
| DRA | 63 * |
| FREN | 3 |
| FREN | 4 |
| HIST | 123 |
| HIST | 147 |
| SOC | 55 |

**Cluster #5**
n = 11

| | |
|---|---|
| ARTH | 186 |
| BIO | 1 |
| BIO | 2 |
| CHEM | 13 |
| CHEM | 14 |
| CHEM | 15 |
| CHEM | 16 |
| EDUC | 135 |
| ENG | 102 |
| HMS | 40 |
| LET | 11 |

**Cluster #4**
n = 15

| | |
|---|---|
| ARTH | 18 * |
| ARTS | 5 * |
| EDUC | 101 |
| EDUC | 134 |
| EDUC | 146 |
| ENG | 55 |
| ENG | 61X |
| FREN | 1 |
| FREN | 10 |
| FREN | 9 |
| HMS | 12D |
| LET | 12 |
| PSYC | 157 |
| PSYC | 79 |
| PSYC | 80 |

**Cluster #6**
n = 8

| | |
|---|---|
| ARTH | 19 |
| BIO | 153 |
| ENG | 85 |
| ENG | 88 |
| HIST | 11 |
| HIST | 148 |
| HIST | 156 |
| HMS | 11 |

**Cluster #7**
n = 2

| | |
|---|---|
| ARTS | 91 |
| HMS | 102 |

**Cluster #10**
n = 2

| | |
|---|---|
| ENG | 101 * |
| MUS | 133 |

## High Neg. Mean Residuals on Analytic Reasoning

**Cluster #4**
n = 15

| | |
|---|---|
| ARTH | 18 * |
| ARTS | 5 * |
| EDUC | 101 |
| EDUC | 134 |
| EDUC | 146 |
| ENG | 55 |
| ENG | 61X |
| FREN | 1 |
| FREN | 10 |
| FREN | 9 |
| HMS | 12D |
| LET | 12 |
| PSYC | 157 |
| PSYC | 79 |
| PSYC | 80 |

**Cluster #8**
n = 16

| | |
|---|---|
| BIO | 37 |
| DRA | 25 |
| EDUC | 136 |
| ENG | 115 |
| ETHS | 130 |
| FREN | 2 |
| FREN | 57 |
| HIST | 121 |
| HMS | 16A |
| M&CS | 62 |
| MUS | 27B |
| PE | 12B* |
| PHYS | 10 |
| SOC | 116 |
| SOC | 94 |
| SOSC | 64 |

## High Neg. Mean Residuals on Logical Reasoning

**Cluster #13**
n = 1

| | |
|---|---|
| GOVT | 17 |

"*" following a course indicates a course misclassified according to the discriminant analysis of course clusters.

Figure 12f. Coursework clusters associated with high negative mean residuals.

**High Neg. Mean Residuals on Quantitative Comparisons**

| Cluster #1 n = 44 | Cluster #5 n = 11 |
|---|---|
| ANTH 57 | ARTH 186 |
| BIO 135 | BIO 1 |
| BIO 191 | BIO 2 |
| BIO 63 | CHEM 13 |
| CHEM 101 | CHEM 14 |
| CHEM 102 | CHEM 15 |
| CHEM 103 | CHEM 16 |
| CHEM 104 | EDUC 135 |
| CHEM 167 | ENG 102 |
| CHEM 4 * | HMS 40 |
| ECON 116 | LET 11 |
| ECON 135 | |
| ECON 136 | |
| ECON 142 | **Cluster #13** n = 1 |
| ECON 155 | |
| ECON 163 | |
| ECON 190 | GOVT 17 |
| ECON 49 | |
| ECON 51 | |
| ECON 52 | |
| ECON 73 | |
| ECON 81 | |
| ENG 62 | |
| GOVT 101 | |
| GOVT 86 | |
| GOVT 90 | |
| HISP 1 | |
| HISP 2 | |
| HMS 9 | |
| M&CS 3 | |
| M&CS 47 | |
| M&CS 48 | |
| M&CS 5 | |
| M&CS 63 | |
| M&CS 64 | |
| PE 102 | |
| PE 12A | |

**High Neg. Mean Residuals on Regular Mathematics**

| Cluster #10 n = 2 |
|---|
| ENG 101 * |
| MUS 133 |

**High Neg. Mean Residuals on Data Interpretation**

| Cluster #7 n = 2 |
|---|
| ARTS 91 |
| HMS 102 |

| Cluster #8 n = 16 |
|---|
| BIO 37 |
| DRA 25 |
| EDUC 136 |
| ENG 115 |
| ETHS 130 |
| FREN 2 |
| FREN 57 |
| HIST 121 |
| HMS 16A |
| M&CS 62 |
| MUS 27B |
| PE 12B* |
| PHYS 10 |
| SOC 116 |
| SOC 94 |
| SOSC 64 |

"*" following a course indicates a course misclassified according to the discriminant analysis of course clusters.

Observations about the cluster relationships

The coursework patterns resulting from the cluster analysis do not produce clear distinctions according to academic department or major. This was true in all samples examined. Departments and programs establish different courses to accomplish varying purposes. It is not surprising to find a "history of ..." course in most fields of study. Likewise, it is not unusual to find a "research methods" course in most disciplines. It should be remembered as well that many of the courses established by a department, program or division are designed primarily to impact specialized knowledge, skills and abilities rather than those associated with general learning. So the coursework associated with student score gains in a given type of general learning does cross disciplinary lines.

In College 3 courses numbered 0 to 99 are lower division courses; courses numbered 100 and above are intended for junior and senior students. Several of the clusters have equal proportions of lower division and upper division coursework. Those clusters associated with Analytic Reasoning draw heavily from upper division coursework in Chemistry (CHEM), Economics (ECON), Math and Computer Science (M&CS). The convention of limiting the stated general education curriculum to lower division coursework may not be warranted.

Different colleges emphasize or provide a different balance to the general learning that occurs. It is clear in College 3 that gains in Analytic Reasoning and Quantitative Comparisons predominate. While the regression analysis evinced large gains in Data Interpretation in College 3, the relationship between these gains and specific coursework patterns was not as apparent. Similar observations were found in the cluster analysis of other college samples. While large amounts of unexplained variance on one or more GRE item-types may appear through regression techniques, the cluster analytic procedure with subsequent secondary validation through discriminant analysis provides a means of associating the formal curriculum with gains in specific learning. Where such relationships are not apparent, student score gains also may be attributable factors not addressed in the analysis: the reliability of the measures themselves, the background of the students enrolling in the courses, the extracurricular environmental variables.

In our analysis of the clusters of coursework in the four colleges and universities examined here, the following conclusions proved to be consistently true:

1. Development of general learned abilities does not have an exact one-to-one relationship with departmental categories. All quantitative reasoning development does not occur exclusively in Mathematics classes. Consequently, simple counts of the number of credits or courses a student has taken in a particular subject may not be a reliable proxy of general learning in the attendant subject area. Quantitative skills, for example, many be developed in a variety of subject areas.

2. Development of general learned abilities is not confined to the lower division. This finding was consistent throughout the four colleges examined. General education requirements of colleges should be re-examined in light of student gains in general learned abilities. Coursework that students who showed significant gains took should be examined, evaluated and incorporated into the general education sequence of the college.

3. There is little 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 student share a common learning experience at a college cannot be readily determined.

The relationships established through the cluster analytic model are associational, not causal. Once a set of courses has been linked to score gains in a specific learned abilities, 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 is 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.

## Conclusions

We began this narrative with the examination of transcripts and assessment test scores from samples of graduating seniors at four colleges and universities in the 1987-88 academic year. From this data, we explored the merits of using incoming student test scores (in this case, SAT scores) as a means for monitoring students learning prior to college. We compared the conclusions about undergraduate general learning that might logically be drawn from the examination of simple GRE mean scores with those that might result from analysis of GRE residuals, once the effect of pre-college learning had been removed. In performing this comparison, we also noted the apparently greater diagnostic and explanatory ability of GRE item-type scores over the aggregate GRE sub-scores which are conventionally reported to colleges. Given the way that many standardized precollege and post-baccalaureate tests are constructed, there appeared to be consequential value in the use of item-types in existing tests as part of a constellation of measures of general learned ability.

We find redeeming value to these standardized tests for a variety of reasons:

1. They provide accessible precollege and postcollege measures without superimposing additional testing on students;
2. They offer established measures with known relationships which clearly measure both collegiate and precollegiate learning;

3. The item-types of these tests show some promise for use as diagnostic indicators of specific learned abilities of students;

4. The known "higher order reasoning" and "critical thinking" measures evince strong correlations with the SAT, the GRE or both, calling into question the efficacy of alternate instruments in measuring general learned abilities.

What is needed to explore the use of standardized measures, such as the GRE, SAT and ACT for assessment purposes is to first device a rationale and mechanism for the testing services to provide colleges item-type data on an optional basis. Second, a longitudinal study of 3 to 5 years is needed to determine which among the item-types has the statistical independence, reliability and construct validity to be incorporated as part of an assessment plan.

In this study we looked not only at test scores but also transcripts. We found them to be useful, usable, non-obtrusive measures of what students chose to constitute their undergraduate education. We drew samples from 10 to 48 percent of the graduating seniors at four separate colleges and universities. It appeared from our analysis that a sample size of at least 10 percent was required in order to have more of the general education coursework adequately represented on the student transcripts. Furthermore, we found that the extent to which students at a particular college held a common intellectual experience (by enrolling in similar coursework patterns) was not a function of institutional size. The percent of unduplicated coursework taken by 5 or more students in any one of the four colleges ranged from 15 to 36 percent. The balance of these students coursework was unique to their program of study, major, minor and electives. At these institutions, the general education requirements comprised from one-third to one-half of the coursework required for graduation. This data suggested that it makes little sense to attempt to make generalizations about the quality of the undergraduate educational experience at these institutions; the transcript analysis suggested that students have far too diverse an educational experience to make meaningful statements about the undergraduate curriculum as a whole.

In contrast, the DCP cluster analytic model provides a consistently reliable means for associating segments of the college curriculum which are associated with a particular group of students, with a particular type of learning (i.e., criterion measure) or both. While our student samples were purposely drawn at random, students could be stratified by incoming ability, thereby producing coursework patterns associated with gains among high or low ability students. Similarly, the criterion measures could include measures of critical thinking, writing analyses, assessments of goal clarity or learning with the disciplines. These measures need not be chosen to the exclusion of the traditional standardized measures of cognitive development, but could be added to them to form the constellation of measures which seems to most accurately describe the educational experience of the students at a given college or set of colleges. To incorporate a variety of different measures into the Cluster Analytic Model, one would only need to statistically standardize student scores prior to performing the cluster analysis in order to insure that no one measure would predominate by virtue of the scale of evaluation it used. All measures would need to have a pre-college assessment counterpart, so that learning prior to college could be controlled and so that the assessment endeavor measures learning rather than admissions selectivity.

## XIII. Bibliography


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.

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

Andrews, Hans A. (1980). The community college catalog and career decision making. Community College Frontiers, 9(1), 11-17.

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.

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

Axelrod, J. (1973). The university teacher as artist. San Francisco: Jossey-Bass.

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

Beattie, C.M. (1984). Behavorial, curricular and demographic variables affecting university students' change in degree major programs. Tallahassee, FL: Unpublished doctoral dissertation, Florida State University, 1984.

Becan-McBride, K. (1980, May). Status of requirements for medical technology curricula. Paper presented at the Annual Meeting of the American Society Medical Technology. (ERIC Document Reproduction Service No. ED 197 657).

Beeken, L.A. (1982). The general education component of the curriculum through transcript analysis at three Virginia Community Colleges. Blacksburg, VA: Unpublished 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., Gould, R.A., & Greenberg, E.M. (1981). Designing Undergraduate Education. San Francisco: Jossey-Bass.

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

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

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

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.

Bloom, B.S., ed., et al. (1956). Taxonomy of educational objectives: Handbook I, cognitive domain. New York: D. McKay.

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. Fourth edition. 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.

Boyer, C.M., & McGuiness, A.C., Jr. (1986). State intiatives to improve undergraduate education: ECS survey highlights. AAHE Bulletin, 38 (6), 3-7.

Boysen, A. R. (1979). Technical communications curriculum for undergraduate colleges of engineering: An institutional analysis based upon selected college catalogues for the 1973-1974 academic year. Dissertation Abstracts International, 40, 1236-A. (University Microfilms No. 7919032).

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

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

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). A multivariate analysis for evaluating distance higher education (Doctoral dissertation, The Pennsylvania State University, 1985). Dissertation Abstracts International, 47, 95A.

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, Colorado: Westview Press.

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

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

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

Cronbach, L.J. (1982). Designing evaluations of educational and social programs. San Francisco: Jossey-Bass.

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). Undergradute curriculum trends. Washington, DC: American Council on Education, 1969.

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.

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 presented 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.

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.

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. (Doctoral dissertation, Iowa Stat 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.

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

Kolb, D. (1976a). The Learning Style Inventory: Technical Manaual. Boston: McBer.

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

Kolb, D., & 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 of 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 re lutions. Chicago: University of Chicago Press.

Lenning, O., Munday, L., & Maxey, J. (1969). Student growth during the first two years of college. College and University, 44, 145-53.

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

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

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.

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

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).

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, May/June). Comparative and predictive analyses of black and white students' college achievement and experiences. Journal of Higher Education, 57 (3), 289-318.

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. 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.

Paykel, E.S. (1975). Nonmetric grouping: Clusters and cligques. 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.

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., & J.K. Wadley. (1976, August). 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).

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.

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

Ratcliff, J.L. (1988a). The effect of differential coursework patterns on general learned abilities of college students: Conceptual framework, research design and methodology. Task #5 Report. U.S. Department of Education, Office of Educational Research and Improvement, Contract No. OERI-R-86-0016, Iowa State University, March 1988.

Ratcliff, J.L. (1988b). 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. (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. (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.

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. (1980). Persistence, Performance, and Degree Achievement of Arizona Community College Transfers in Arizona Public Universities. Tempe, AZ: Arizona State University, Dept. of Higher and Adult Education. ERIC Document Reproduction Service No. 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.

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

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

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. Sternberg (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.

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

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. Vermilye (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.

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.