DOCUMENT RESUME

ED 357 700 HE 026 448

AUTHOR Ratcliff, James L.; And Others
TITLE Handbook on Linking Assessment & General Education.
INSTITUTION National Center on Postsecondary Teaching, Learning, and Assessment, University Park, PA.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Jun 92
CONTRACT R117G10037
NOTE 122p.; For related documents, see HE 026 449-551.
PUB TYPE Reports - General (140) -- Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Academic Achievement; *Cluster Analysis; Curriculum Development; Educational Improvement; Educational Planning; Evaluation Methods; *General Education; Guidelines; Higher Education; Learning Experience; *Models; Outcomes of Education; Program Development; *Program Implementation; *Student Evaluation; Undergraduate Study
IDENTIFIERS *Coursework Cluster Analysis Model

ABSTRACT

This handbook presents a process for the assessment of college curricula and student learning in general education and liberal learning. The key concepts of the Coursework Cluster Analysis Model (CCAM) are described and explained. Some of the general guidelines on interpreting the results of CCAM analyses are presented as well as key decisions regarding reform of the curriculum, although revision of the assessment program and enhancement of the student advising system will have to be made on each campus employing the CCAM model. Specific chapters discuss CCAM and the assessment of student learning, the research that went into developing CCAM, the considerations for planning for the assessment of student learning, a general overview of the methods and procedures in using CCAM, determining student learned abilities, and the detailed procedures for conducting the coursework patterns analysis. Appendices provide studies of the CCAM analysis and the job control language required to perform the CCAM procedures. Contains 50 references. (GLR)

ED357700

Handbook on Linking Assessment & General Education

by

James L. Ratcliff
Elizabeth A. Jones
Steven Hoffman

June 1992

A handbook developed and distributed by the National Center on Postsecondary Teaching, Learning, and Assessment (NCTLA). NCTLA is funded by the U.S. Department of Educational Research and Improvement (OERI), Grant R117G10037. The opinions herein do not necessarily reflect the position or policy of OERI, and no official endorsement should be inferred.

HE 026 448

NATIONAL CENTER ON POSTSECONDARY
Teaching, Learning, & Assessment

The Pennsylvania State University, 403 South Allen Street, Suite 104, University Park, PA 16801-5202
(814) 865-5917 FAX: (814) 865-3638 BITNET: NCTLA@PSUVM

Project No.: R117G10037
CFDA No.: 84.117G

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

CHAPTER ONE
THE COURSEWORK CLUSTER ANALYSIS MODEL AND THE ASSESSMENT OF STUDENT LEARNING

What It Is, What It Does, And What It Doesn't Do

The value of assessment of student learning is in its ability to point the way to more effective educational programs and ultimately, to improved student performance in college. For assessments to be of value in improving student learning, they must not simply tell us how well students have learned, but they must link that learning to students' educational experiences. Assessments of student learning in general education are particularly difficult to marshall in comprehensive research universities, doctoral-granting institutions, and comprehensive colleges. Here the curricula and the programs are diverse, often consisting of thousands of courses loosely organized into distributional plans of general education.

Students who take different coursework learn different content, cognitive skills, values and attitudes. Student learning varies greatly in complex institutions of higher education because of their broad arrays of curricular offerings. Critical to the success of a general education for students in these institutions is some means for recognizing curricular diversity and its effects. Thus, the more complex the curricular offerings, the greater is the challenge it is to determine the relationship between coursework taken and learning achieved.

Most faculty and administrators are committed to improving the quality of undergraduate education. And, to make improvements, it is necessary to know what students learn in order to decide what ideally they should learn. Assessment plans and programs can monitor institutional performance relative to student learning. Eighty percent of the nation's postsecondary institutions report using some outcomes assessment activities to evaluate program quality and educational effectiveness (El-Khawas, 1989).

However, most institutions are in the beginning stages of planning and designing assessment activities. The American Council on Education (1991) reported that only 30 percent of the nation's two and four-year colleges and universities operate comprehensive student assessment programs. An additional 60 percent reported that they planned to establish such programs in the future (ACE, 1991). At the same time, most institutions are attempting to reform general education (Gaff, 1992). Unfortunately, rarely have assessment initiatives and curricular reform efforts informed one another. While the evidence suggests that most colleges and universities are trying to improve student learning, the results of their efforts have yet to materialize (Astin, 1991; Eaton, 1991). The Coursework Cluster Analysis Model (CCAM) provides a way for faculty and administrators to make more substantive links between what students study in college and what they learn.

The model and method of analysis defined in this handbook permits a college or university to do several things:

- determine which assessment measures best describe the kinds of learning that take place among students at their institution;
- determine which parts of the curriculum are currently not monitored or described by the present assessment methods and measures;
- determine which patterns of coursework are associated with which kinds of learning and with which groups of students;
- determine the extent to which transfer students benefit from the same or different general education coursework from that taken by students who began their baccalaureate program at the same institution;
- determine the extent to which a core curriculum or a distributional requirement produces the greatest gains in learning among different groups of undergraduates at the same institution;
- determine which course sequences contribute to general education and liberal learning and which do not.

The CCAM has some limitations as well. It is intended for:

- assessment of general education and liberal learning, not learning within the major;

- those institutions that have a distribution plan of general education wherein students have a fairly wide range of curricular choices from which to fulfill the requirements for their baccalaureate;

- identification of coursework *associated* with improvement in student learning in general education and liberal learning. It does not tell us that coursework *caused* that learning. Subsequent research and analysis is required to determine what factors contributed to that learning.

The Missing Link: Curricular Content and Student Learning

The 1980's were a decade of examination of the state and quality of educational programs. National reports urged faculty and academic leaders to improve baccalaureate programs. The Study Group on the Conditions of Excellence in American Higher Education, formed under the U.S. Department of Education, urged colleges to provide students clear academic direction, standards, and values. It urged researchers to use college student assessment information and to explore the use of student transcripts as resources in understanding more about what subjects students study in college and what they learn. The procedures and techniques of student and curricular assessment described in this handbook are a direct outcome of those recommendations. Beginning in 1985, we developed specific procedures to determine the gains in student learning that were directly attributable to enrollment in different patterns of undergraduate coursework.

In February 1985 the American Association of Colleges (AAC) issued a report, ***Integrity in the College Curriculum***, which drew the decidely negative conclusion that undergraduate education was in a state of crisis and disarray. The report attacked "marketplace" oriented curriculum based solely on student choice, asking "Is the curriculum an invitation to philosophic and intellectual growth or a quick exposure to the skills of a particular vocation?" (p. 2). The report called on colleges and universities to live up to their stated goals for general education and liberal learning by providing a coherent curriculum. For AAC, a coherent curriculum at least insisted on inquiry, literacy, understanding numerical data, historical consciousness, science, values, art, international and multi-cultural experiences, and study of some discipline in depth (Eaton, 1991).

At least three studies have tried to determine what improvements in college curriculum have been accomplished since 1985. Robert Zemsky (1989) examined 35,000 student transcripts from 30 colleges and universities to determine the shape and substance of the curriculum they had encountered. Zemsky found a continued lack of curricular structure and coherence, that students' enrollment in science and mathematics was quite limited, and that the humanities lacked sequential, developmental patterns of learning. Lynne V. Cheney, analyzed humanities enrollments in colleges and universities to determine if there had been a fundamental change in baccalaureate programs between 1983 and 1989. She found little, if any change in undergraduate degree requirements. She lamented,

> It is possible to graduate now, as it was five years ago, from more than 80 percent of our institutions of higher education without taking a course in American history. In 1988-89, it is possible to earn a bachelor's degree from:
>
> o 37 percent of the nation's colleges and universities without taking any course in history;
>
> o 45 percent without taking a course in American or English literature;
>
> o 62 percent without taking a course in philosophy;
>
> o 77 percent without studying a foreign language (1988, p. 5).

Not only was their little evidence of increased structure and rigor to the curriculum during this time period, there was also evidence that the curriculum was not having much impact on student learning. Alexander Astin, in a national study of student transcripts, general education requirements, and student test scores and self-reports, found no relationship between any general education curricular structure and improvement in student learning (Astin, 1991, AGLS meeting).

While there were strident calls to improve undergraduate education from these national reports and studies, colleges and universities did not remain idle. During the past decade more than 90 percent of colleges and universities engaged in some kind of revision or reform of their undergraduate curriculum (Gaff, 1989). The American Council on Education repeatedly reported in ***Campus Trends*** that most colleges and universities were engaged in curricular reform. This has led Judith Eaton (1991) to raise some rather uncomfortable questions about this flurry of activity:

> From a negative point of view, one can point to little in the way of completed curricular modifications or, more important, changes in student performance that ... emerged ... as the 1980's ended. Worse, one might view the decade ... as an essentially unimportant ten-year saga during which the higher-education community continued an apparently endless and unproductive dialogue with itself on academic issues as opposed to engaging in constructive action (p. 61). Did institutional descriptions of academic reform fail to focus on those intended to benefit but, instead, confused expectations of student performance with descriptions of faculty involvement? (p. 63).

We have yet to make a meaningful connection between undergraduate curricular content and improved student learning. The increased national attention given to improved student performance and stronger academic direction, standards and values call upon us to make more substantive links between what students study in college and gains in their learning. Why are faculty and administrators focusing more attention on the assessment of student outcomes?

The Impetus for Assessment

A variety of both external and internal factors are compelling institutions to not only consider assessment but to formalize plans and take specific actions to measure the educational impact of an institution on its students. One group of external factors involve a variety of state initiatives. Dissatisfaction with student learning has lead an increasing number of states to expect colleges and universities to implement student assessment programs. Earlier state policies toward assessment took a decentralized approach, allowing institutions to develop their own systems of assessment. However, state policy makers are increasingly dissatisfied with assessment programs that do not improve student learning. The result has been new state proposals for common student outcomes testing (Ewell, 1991). Some states have adopted formal assessment requirements while many other states are moving in this direction. Every student in Florida who is preparing to receive an associate's degree from a two-year institution or who plans to become a junior in a four-year institution is required by the state to take the College Level Academic Skills test (CLAST). Since 1979, Tennessee has based part of its public college and university funding on student assessment results. Institutions in Tennessee test seniors in general education and in their chosen majors, survey alumni, and use the results of assessment activities to guide improvements at the institutions (Banta, 1989).

Another set of external factors are the accrediting organizations. Most of the six regional accrediting associations have begun to incorporate outcomes assessment as a criterion for institutional approval. The North Central Association of Colleges and Universities has conducted regional seminars on assessment and prepared a workbook to aid in the evaluation of institutional effectiveness and student achievement. In addition, accrediting bodies that approve programs in the disciplines are beginning to include outcomes assessment in their criteria for approval.

Due to these external factors, institutions developed and implemented assessment programs often to provide accountability. However, there are internal factors which have encouraged institutions to undertake assessment activities for the sake of academic improvement. The information gathered from assessments can help reform the curriculum; strengthen academic programs and student services; and consequently, increase student satisfaction and enhance student recruitment and long-term retention. Using the information from assessment activities, faculty can give specific attention to the need for self-improvement in teaching and evaluating students in their own individual courses. The model described in this handbook is focused on assessment for the purpose of academic improvement.

Development of the Coursework Cluster Analysis Model

Assessments describe and document the nature and extent of learning that has occurred. They cannot tell us, however, which courses most consistently produce gains in learning for specific groups of students over time at particular institutions. Such information would be extremely useful. Knowing the degree to which different courses contribute to different learning outcomes would provide a college or university with an empirical basis for curriculum review. Knowledge of such links between coursework and learning could serve as a powerful source of information which would complement faculty wisdom, student evaluation, and other means of appraising the extent to which particular sets and sequences of courses have the effect for which they were intended. Such information could also be used to improve academic advising and guidance students receive in making course selections (Ratcliff, 1990a, 1990b, 1990c).

Over the past four years we have developed a model for linking assessments of the general learning of undergraduates with the coursework in which they enrolled (Ratcliff, 1988, 1989, 1990a, 1990b; Ratcliff & Jones, 1990, 1991; Jones & Ratcliff, 1990a; 1990b; 1991). This research has proceeded under the rubric of the Differential Coursework Patterns (DCP) Project, and the model for linking coursework to student assessment has been referred to as the Coursework Cluster Analysis Model (CCAM). Its development and testing was supported first by the Office of Educational Research and Improvement of the U.S. Department of Education. Subsequent qualitative validity studies of the GRE item-types, trend analyses of coursework patterns, and studies of the applicability of the model to curricular reform, assessment program development and academic advising has been supported by the Exxon Educational Foundation. The CCAM has been tested at six institutions: Stanford and Georgia State Universities, and Clayton State, Evergreen State, Mills and Ithaca Colleges. In addition, CCAM has been applied to student reports of enrollment patterns and ACT COMP scores at the University of Tennessee--Knoxville (Pike & Phillippi, 1989).

In the most typical applications, assessment instruments were administered to graduating seniors. The results of these post-tests were compared with the results of corresponding pre-tests of the same students. Such well-known standardized instruments were used: the SAT, GRE, ACT and ACT COMP examinations, as well as the Kolb Learning Styles Inventory and locally-constructed measures of student-perceived course difficulty. A great strength of the Model and an asset that seems to enhance its acceptability to faculty is that it is not dependent on instruments supplied by external vendors. It can use a variety of locally-developed instruments, tailored to particular needs and extensively employing local judgment. A college, for instance, might administer its own essay examinations to freshmen and seniors, and its own faculty might grade them holistically; so long as the final evaluation, and/or its subparts, can be translated to a numeric scale, this instrument would be entirely adequate for the purpose of the Coursework Cluster Analysis Model.

A common stumbling block in the development of an assessment program is that of what form of test or assessment information to use. Curricular reviewers, reformers and researchers quickly acknowledge that there is no clear conception of what constitutes general learning. Such recognition emerges regardless of whether it is the college curriculum or the various tests and assessment devices that are being examined. A college that attempts to reach consensus among its constituents on either general education goals or on the "best" measure of general learned abilities will foster heated discussion. The quest for consensus on what should be the common intellectual experience of undergraduates may end in irresolution or, worse, abandonment of the assessment initiative. Instead of searching for the ideal measure of general learning in a college, ***those charged with assessment can better direct their energies toward the selection of a constellation of assessment means and measures that appear to be appropriate criteria for describing one or more dimensions of the general learning goals of the college.***

The Coursework Cluster Analysis Model provides a basis for determining the relative extent to which each measure explains general student learning within a given college environment. If we have nine different assessment measures, for example, we can determine what proportion of the variation in student scores was explained by each measure. This information leads to a decision-point for the academic leader or faculty committee charged with the development and oversight of the assessment program. If a measure of general learning does not explain much of the variation in student scores, one option is to conclude that the measure is inappropriate to the students and the educational program of that particular college or university. In short, it can assist in the discard of that particular form of evaluation as superfluous and unnecessary. An alternative conclusion is that the institution is not devoting sufficient attention to the type of learning measured. Here, an examination of the assessment instrument itself relative to the curriculum is called for. Again, the Coursework Cluster Analysis Model can point to those courses and classes that were associated with gains in student learning on the measure in question.

How This Handbook is Organized and Who It is Designed For

This handbook is organized into six chapters. Each chapter describes critical central concepts and strategies for implementing an assessment program using the Coursework

Cluster Analysis Model. Chapter One describes the Model and benefits of using it in assessing general education and liberal learning. Chapter Two details the design of the Model and defines key concepts. Chapter Three shows how to plan an assessment program including the identification and selection of potential measures, instruments, and tests. Chapter Four outlines the methodology and procedures to be followed using the Model. Chapter Five tells how to determine the nature and extent of student learning in general education and liberal learning. Chapter Six describes how to use discriminant analysis and cluster analysis to determine which sets of courses are linked with improvement in which types of student learning. In each chapter we discuss the central concepts and provide concrete examples where needed in order to give a full picture of the major aspects and attributes of the Model. At the end of each chapter, we provide a summary of the major points.

This handbook will be useful to faculty and administrators interested in conducting an assessment of general education and determining its impact on students at their institutions. Provosts, academic vice-presidents and deans, associate or assistant deans will find general information concerning the advantages, steps, and limitations of the Coursework Cluster Analysis Model and alternative ways to measure student learning in Chapters Two, Three and Four.

Institutional researchers or directors of assessment programs can find detailed information on how to link student assessment and transcript curriculum information using the Coursework Cluster Analysis Model in Chapters Five, Six, and in Appendices. This information is often presented in a step by step manner to enable those interested individuals in actually implementing this model to determine what information they need to gather to assess student learning at their institutions. We use concrete examples with specific assessment measures to illustrate the application of the ideas presented in each chapter. In appendices, we describe in even more detail the specific computer commands needed to produce the desired analyses using SPSS.

Parameters of the Handbook

The Coursework Cluster Analysis Model presented in this handbook is designed for use by colleges and universities with distributional general education requirements or with a large array of curricular choices for undergraduates. The handbook will have limited value for those institutions with a core curriculum for general education or a small number of course offerings. In addition, the methods and procedures described assume that institutions have the capability to merge student transcripts with assessment measures.

This handbook and its presentation are also premised on several other assumptions. First, courses are the primary units of learning in college. Second, learning is not merely the sum of all courses, but is actually developmental and cumulative, requiring the identification of combinations and sequences of courses. Third, transcripts are an accurate listing of the enrollment pattern of students. Fourth, most undergraduate courses are basically stable in content and instruction over time and among instructors. Fifth, the effects measured can be generalized to all the formal coursework in which a student enrolled. These are common assumptions underlying the development of the Coursework Cluster Analysis Model. It links the coursework students take with their general learned abilities.

The Model may be used to measure (a) content learning with disciplines, (b) cognitive development in general learned abilities, (c) motivation and attitudes toward learning, and (d) persistence and progress toward degree attainment (Terenzini, 1989). The Model has the most potential and application to measure the cognitive abilities of students.

Cognitive outcomes would be defined by the actual instruments used and what they are attempting to measure. For example, if the Graduate Record Examination was used these cognitive abilities would include Reading Comprehension, Sentence Completion, Antonyms, Analogies, Logical Reasoning, Analytic Reasoning, Data Interpretation, Quantitative Comparisons, and Regular Mathematics. Cognitive domains have also been defined by Bloom (1956) as knowledge, comprehension, application, analysis, synthesis, and evaluation. Bloom's domain categories represent some of the various levels of cognitive functioning which faculty may want to consider when selecting instruments to assess student learning. An important goal is to use a variety of items which assess these levels of abilities. The Coursework Cluster Analysis Model may also be used to assess affective or non-cognitive outcomes such as willingness to participate in class, self-esteem, sense of autonomy, or sensitivity to the needs of others. In Chapter Two, we present a more detailed discussion of potential measures.

Benefits of the Handbook

A number of benefits are anticipated from the implementation of this Coursework Cluster Analysis Model in the assessment of general education. First, the Model can use multiple measures of assessment; thereby allowing for a broader picture of student learning than any one measure can paint. It provides institutions with information regarding the extent of variation in student assessment results that is explained by any one of the measures used. This information can be helpful in a number of ways. Faculty and administrators need not decide on an ideal set of assessment measures. The extent to which such measures may overlap in describing student learning can be identified. The mix of assessment measures appropriate to the goals of the college and the characteristics of the student population can be continuously monitored. When students show small amounts of growth on an indicator of student learning, either the college can develop strategies for improving student learning in the area identified, or discard the measure as inappropriate to the college and its students.

Efforts to assess general education and liberal learning can become quickly bogged down in discussions over which measures, indicators or examinations to use. Faculty feel pressured to commit to a set of measures that may not accurately reflect their vision of the goals of general education. By using multiple measures and by leaving the decision as to which measures to use open to continuous revision and updating, the college or university can proceed to develop a rational, cogent and informative assessment plan. Judith Eaton has written about tensions that emerge over the discussion of the desired outcomes of general education and the desirability of such a contingency approach:

> These tensions emerge when we are either unwilling or unable to commit some defensible approach to general education for fear that our commitment will be found lacking in some way. Waiting around for the ideal general education scenario, however, serves little purpose and harms students even more than a general education effort that possesses some flaws (Eaton, 1991, p. 66).

If a general education innovation holds promise to enhance student learning in some way, then there should be a means to ascertain whether that improvement has occurred or not. Linked analysis of assessment and enrollment data holds the promise of identifying when and, more importantly, under what circumstances improvement have been made to the general education curriculum. The Coursework Cluster Analysis Model provides useful information to the college about the mix of assessment measures that reflects what the students learn and what the college intends to teach them.

The Model can provide concrete useful information about the curriculum that can guide reform efforts. The Coursework Cluster Analysis Model is a tool ideally suited to institutions of higher education with a distributional general education requirement and a wide array of programs, electives and majors. From a catalogue of hundreds or thousands of courses, CCAM can identify those that are taken by students who showed the greatest improvement in learning. For example, if one of the assessment measures a college selects is a test of analytic reasoning, then the CCAM can identify those groups of courses that students took who showed significant improvement in that area of general learning.

In the Differential Coursework Patterns (DCP) Project which lead to the development of the CCAM, we found little relationship between gains in general learning and the formal general education requirements of the colleges and universities we studied.

However, everywhere we looked students who took different coursework learned different things and developed different abilities. There are two lessons from this research. First, what students take in college does have bearing on what they learn. Second, the structure and sequence of general education in the institutions we examined did not produce a profound effect on the types of learning we examined. Our research not only affirms that what we provide in the college curriculum does make a difference, but it follows common sense.

The finding that studying different courses leads to different types of learning is really corollary to a larger, more important research finding. While this was affirmed in our research, it is best described in Ernest Pascarella and Patrick Terenzini's important new book, ***How College Affects Students*** (1991). Here they describe and analyze 20 years of research indicating that ***differences in student learning are far greater within institutions than between them***. Given this finding, it stands to reason that students taking different coursework and having different extracurricular experiences should show differences in subject matter learned, in the type and extent of their general cognitive development, and in differences in values and attitudes toward learning.

This finding that variation in student learning is greater within colleges than between them also means that one intellectual shoe does not fit all freshmen feet. The efficacy of a single set of courses, a core, to fostering the intellectual development of college students can be easily examined using assessment results. Divide a group of graduating seniors into those who entered college at or above the mean of SAT scores and those below the mean. Sort (cluster analyze) the courses these students took by the gains they demonstrated on the assessment measures (Ratcliff & Jones, 1991). If a core curriculum would be a superior arrangement, then both the high ability and low ability students who showed large gains would have taken basically the same coursework. If these two groups would benefit from distinctive different curricular sequences, then those who showed large gains from the low ability group would have taken different courses than those who showed large gains from the high ability group. Assuming that the assessment criteria mirrored the intent of the curriculum, where overlap between the two groups occurred, a core curriculum would be justified; and where there was no overlap, separate curricular sequences for each group would be appropriate means to achieve the general education goals. The DCP Project research on this problem does not argue for a single core curriculum. In each institution with a distributional general education plan we examined, our research indicated that the implementation of a unity core curriculum might actually mitigate against student learning, particularly among lower ability students. Therefore, it may be more productive and desirable for a college to examine important student groups by background and ability.

The student population can be subdivided into high ability and low ability students, by gender, race or ethnicity, or by major. Then the Model can identify if the coursework associated with gains in learning among the total group is the same as that for the subgroups. Curriculum planners and curriculum committees can readily use this information. Courses in the general education sequence not associated with gains in student learning can be revised, enhanced or dropped. Courses outside the general education requirements that contribute to gains in student learning can be included in the general education curriculum.

The Model can also produce information that can lead to better academic advising, since it links the coursework students take with their improvement in learning. Students can choose from lists of courses taken by others with similar backgrounds and abilities--others who showed gains in performance and learning. It takes advising beyond the mere listing of formal degree requirements. As more data are amassed, greater and greater precision is generated in the linking of coursework and student learning. The Model may even be amenable to the development of a microcomputer-based advising system utilizing a relational database of prior students coursetaking patterns and assessment results. Such a computer-based advising system would yield an array of effective coursework tailored to the abilities and interests of individual students and within the parameters of institutional degree requirements.

Summary

This handbook is designed to provide a process or procedure for individuals to assess student learning in general education and liberal learning. In the next chapter, the Coursework Cluster Analysis Model and key concepts will be described and explained in greater detail. It is a tool which links the coursework students take with their improvement in learning. This information is not an end in itself rather it is a means to an end. Curriculum reform can proceed guided by concrete information on what coursework helps which students the most. Assessment measures may be continually updated and revised according to changes in curricular purpose or student achievement. Students can get information on courses that have been beneficial to others like them. Teaching and learning can be advanced. Assessments of student learning are complex matters but by using multiple measures in this Model information can be gained concerning what courses help students to gain in particular areas of cognitive development.

While this handbook presents a method to assess student learning and to gain valuable information on student outcomes, the college community must make many decisions throughout this process. These decisions include what are the purposes of assessment, who is to be assessed, what will be assessed and how will it be assessed. This handbook primarily tells how student learning can be assessed by using the Coursework Cluster Analysis Model but it leaves to academic leaders and curriculum committees the key decisions about what assessments are best given their college environment. While this handbook provides some general guidelines on interpreting the results of the CCAM analyses, key decisions regarding reform of the curriculum, revision of the assessment program and enhancement of the student advising system will have to be made in each campus employing the CCAM model. The primary purpose of this Model and its application is to improve and promote student learning. Through the enhancement of student growth, general education and liberal learning also increase their effectiveness. This handbook briefly presents information concerning who to assess as well as what instruments are available to assess student cognitive abilities. However, there is a wealth of literature which provides a more comprehensive presentation in these areas. We provide information on assessment primarily to exemplify and to describe best how to use the CCAM to improve student learning in general education and liberal learning.

CHAPTER TWO
THE RESEARCH BEHIND
THE COURSEWORK CLUSTER ANALYSIS MODEL

Linking Coursework and Student Learning

Before you plan your assessment program and before you begin to link your curriculum to what you learn from that assessment, you need to establish some general agreement on some concepts that may be otherwise taken for granted. These are deceptively simple questions, such as:

- What constitutes student learning?
- What is a course sequence?
- What is the undergraduate curriculum?
- What is the difference between curriculum intents and curriculum outcomes?

In this chapter, key definitions are provided for the major concepts which form the design of the Coursework Cluster Analysis Model. Since the Model does not rely on a particular theory of student learning, the areas of general education and liberal learning, student achievement, coursework patterns, and the data sources need to be clarified. The conceptual framework for analyzing the coursework patterns is reviewed as well as limitations in the analysis of these patterns. The issue of the representativeness of the curriculum found on the student transcripts is also discussed. This background information sets the context for Chapter Four which deals with an overview of the methodology and procedures used in the Coursework Cluster Analysis Model.

What Constitutes Student Achievement?

A question encountered in determining what constitutes general learned abilities is that of what composes the ***gains*** resulting from a college education. Simply measuring how graduating seniors perform on a series of tests is not a sufficient basis for generalizations about the effect of college on student achievement. The assessment of student outcomes is heavily affected 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 post-college tests measure are attributable to student learning prior to college.

What Constitutes a Coursework Pattern?

The prevalent way to view the college curriculum is by its intentions, rather than by its results (Warren, 1975). Since measuring the effects of the curriculum is problematic, it is not surprising that many studies presume rather than test the effect of different patterns of coursework of student learning.

Here we begin from the opposite position. Instead of inventing a new curriculum and then determining whether student learning surpasses that induced by the older one, we determine what parts of the existing curriculum are most associated with students who show the most learning. It operationalizes the research showing that variation in learning is greater within institutions than between them, and that different coursework is associated with different types and levels of learning (Ratcliff, 1990; Pascarella & Terenzini, 1990). Thus, we want to look at all courses and see which are associated with attainment of general education goals. This is better than looking only at designated general education coursework and asking how it is working because it provides an empirical basis for the inclusion or exclusion of specific sets of courses. It identifies developmental relationships among courses and it allows for variation in student learning and student course selection.

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 & Grasha, 1983). The medieval university curriculum was organized according to combinations and sequences of courses as well as individual subjects (Rudolph, 1977); the seven liberal arts were sequenced into the prerequisite subjects of the ***quadrivium*** (arithmetic, geometry, astronomy, and music) and the higher order subjects, the ***trivium*** (logic, grammar, and rhetoric). Together, the quadrivium and trivium provided an individual with the general learned abilities needed to study the three philosophies of Aristotle: natural philosophy (physics), moral philosophy (ethics), and mental philosophy (metaphysics). These combinations and sequences of coursework have been generalized more recently into concepts of breadth and depth as criteria by which to describe higher education curricula (Blackburn et al., 1976).

The notion that combinations of concurrent coursework and developmental sequences of coursework leads to improved student learning dates back to the medieval university. Current research and learning theory affirms its value as well. Perry (1968) for example, saw that development "consists of an orderly progression of cognition in which more complex forms are created by the differentiation and reintegration of earlier, simpler forms" (p. 44).

The value oı curricular substance and sequence appear in various ways in the curriculum such as the formulation of a core curricula. The merit of course sequencing is implicit in the four levels of study (freshman, sophomore, junior and senior years). Colleges assign course numbers to indicate when they should be taken by students and assign course prerequisites. To fully assess the impact of the curriculum on student learning, the additive, substantiative and sequential characteristics of student course-taking need to be examined. These notions of what ought to be taught and what students ought to learn presumably represent the philosophical and educational aims of the particular college.

Nevertheless, a distinction should be made between those patterns of coursework *intended* to fulfill undergraduate program and degree requirements and those patterns of coursework which students actually choose (Boyer & Ahlgren, 1981, 1982, 1987; Warren, 1975). 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. They alone do not tell us anything about the effectiveness of the curriculum.

In a college curriculum, a single course may be the smallest unit of analysis. A pattern of courses is a design resulting from their relationship to one another (Romesburg, 1984). A cluster of courses is a set of one or more courses found to be similar according to a given set of attributes. In the DCP Project, courses were grouped according to the extent of gains (or losses) in general learning of the students enrolled. Thus, for the purposes of the Coursework Cluster Analysis Model, a ***cluster of courses*** is a pattern of coursework with an empirically derived set of relationships. Stated another way, a cluster of courses is a pattern based on student learning, not necessarily on what the college thinks is good for students to take to fulfill the general education requirements. This distinction between intent and effect is most important. It sets apart the college committed to assessment and a results-oriented curriculum.

Student Transcripts as a Data Source

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

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

Prather and associates (1976) used transcript analysis to study undergraduate grading practices at Georgia State University. They investigated differences in grading patterns by major fields of study while controlling for such factors as scholastic aptitude, demographic background, course types, and longitudinal trends. They found that major field was strongly associated with the grades students received. This research showed that different parts of the curriculum have different grading standards. It argued against using the GPA as a proxy measure of student learning since grades varied among subjects. Prather and associates, however, used an electronic database of student transcript information to examine an institutional cohort; use of electronic databases of records enabled the researchers to examine larger samples of student records, thereby permitting analysis of larger sections of the curriculum.

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

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

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

Transcripts are a useful, valid and reliable source of information on student course-taking behavior. They provide evidence of the combination, sequence and performance of students in the patterns of courses in which they enroll. They are unobtrusive too. While some studies have been limited in scope because it takes so long to examine individual transcripts by hand, there is growing evidence that such records, stored on a college or university computer, can be readily used to examine the course-taking behavior of a whole class, cohort or population of students. The CCAM uses transcripts in precisely this manner. Transcripts maintained on an electronic database can be merged with student assessment scores to link the curriculum to student learning.

A Conceptual Framework for Analyzing Coursework Patterns

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

We begin the problem of linking coursework to student learning knowing that what the student's assessment results were and what the courses were that the students took. We know that a student may have taken 35-45 courses in a baccalaureate program, but we do not know the contribution of any one course to gains in learning that a student may have demonstrated.

In order to determine the *improvement* in learning a student demonstrated, we first must determine the proportion of this student's assessment score that is explained by the student's entering ability assessment. Afterall, a student enters college with 12 years of formal education and many life experiences which contribute to the student's ability to complete the outcomes assessment accurately.

There are several means of controlling for the student's entering ability in assessing this student's improvement in learning. The method we have used in the DCP project has been to regress the students' postcollege scores on corresponding precollege scores. The residuals (that part of the score not explained by the student's entering ability) then becomes the proxy measure of this student's learning for each assessment measure used. Thus, the first step in the analysis is to determine how much of a student's general learning is attributable to college.

Once the net effect of the college years is separated from the other educational achievement of students, then the task of linking collegiate learning with the curriculum can begin. The net effect of collegiate learning for every student enrolled in a particular course is tagged to that course. Thus, it twenty students enrolled in a math course (as evidenced by their transcripts), then the gains in student learning for each assessment are tagged to that math course. This is done by computer for each and every course taken by students in the assessment program.

Next courses are sorted and grouped into clusters according to their similarity of net effect. Coursework taken by students showing larger gains in analytic reasoning may appear separately or together with the coursework associated with improvement in reading comprehension, depending upon the relationship of those assessment measures in the student's learning experience. Coursework linked to the net effect gains in learning separates from coursework taken by students who showed little or no improvement in one or more of the assessment criteria. Coursework not linked to gains may be valuable for other purposes or reasons (learning in the major, etc.), but it is disclosed as having no direct relationship to improvement according to the general education assessment criteria. Thus, the Coursework Cluster Analysis Model sets the stage for further exploration of the specific role coursework plays in the general education curricula.

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

Likewise, the effect of a particular course may depend on the other courses in which the student is concurrently enrolled. Richardson et al. (1982) and Roueche and Snow (1977) noted that students may be advised to enroll in elementary writing or mathematics courses concurrently with other courses requiring the basic skills these elementary courses teach. Under such practices, the student may have much less chance to succeed in college. Particular combinations of courses may produce specific positive or negative effects (Bergquist, et al. 1981; Rudolph, 1977; Veysey, 1973). Since the CCAM can be used to identify the term in which students enrolled in a particular course and courses taken concurrently, the effect of such patterns can be examined.

A student may choose a particular course at a particular time in his/her program of study for any number of reasons. A poor grade in "Trigonometry" may cause Student Y to select a remedial mathematics course over "Introduction to Calculus". Student X, who received a high grade in "Trigonometry", may not enroll the following term in "Introduction to Calculus" because the time it is offered conflicts with that of another course Student X is required to take. Or the Calculus course may be filled when Student X tries to enroll. Many factors shape the combination of courses a student chooses in a given term and the sequence of courses represented across terms in the transcript.

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

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

1. students make course selections in an environment of uncertainty about the consequences of their choices;
2. there are multiple reasons why students enroll in each course;
3. there are multiple options available to the student at each decision-point (term or registration period);
4. student course selections are sequential; there are different decision-points (terms) in which parts of the coursework pattern are chosen, with prior decisions having some bearing on future decisions.

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

According to Pace (1979), one variable in student development is the amount of time and effort invested by the student. This premise, that student involvement in learning advances student achievement, guided the recommendations of the NIE Study Group's Report on Conditions of Excellence in American Higher Education (1984). Not only the kind and quality of cognitive activities in which the student engages, but also the level of effort exerted by the student in understanding and using the knowledge and abilities gained influence the quality of student learning. The student's effort in courses is "impressed" (Pace, 1979) by attitudes of the perceived usefulness of the course and the perceived difficulty of the course. These perceptions influence the kind and quality of student investment in learning. Coyne and Lazarus (1980) found that such investment involved both cognitive and subjective elements, leading to whether the experience is viewed as a challenge or a threat. The perceived difficulty of courses influences student enrollment decisions and thereby contributes to the multiple-stage enrollment decision-making process through which the student compiles his or her particular collection of coursework.

In summary, the literature suggests a number of possible interactions between student and curriculum each time a student makes course selections. The effect of courses on student learning in general education may vary according to the course itself, the time of enrollment in the student's baccalaureate program, the concurrent or sequential relationship to other courses in which the student enrolls, the predominant learning style of the course and of the student, the curricular design of the course, and the risk-taking behavior the student exhibits at each enrollment decision-point. The Coursework Cluster Analysis Model calls first for the identification of student achievement (i.e., the net effect of student score residuals) and secondly for the classification of courses found on student transcripts into patterns according to their associated effects on the student score residual. The model provides a basis for examining the extent to which the empirically-derived patterns of coursework reflect institutional mission and curricular goals, general educational requirements, the values, norms and mode of inquiry represented by the disciplines studied, and the demographic characteristics of the students. The model accomplishes these objectives through the use of cluster analysis, a statistical procedure which has been used throughout the physical and social sciences to derive empirical taxonomies of objects in a variety of settings. Cluster analysis has been infrequently employed in education and is described in greater detail in Chapter 6.

The Relationship Between Student Sample Size And The College Curriculum

"How big a sample of students do I need to get at the kinds of questions CCAM can help us answer?" The answer to this question of sample size is in part determined by the curriculum and in part by the measures selected. A persistent problem in linking the undergraduate curriculum to measures of student learning is the number of courses from which students may choose. As mentioned earlier, students will enroll in 35 to 55 courses to complete their bachelor's degree, although the number varies considerably from student to student. Students select these 35 to 55 courses from a catalog of several thousand courses at a university or several hundred at a smaller college. Linking the effect of sets of courses to the general learning of students therefore becomes complex.

First you identify all the courses appearing on the student transcripts. Next you need to single out those courses that were cross-listed or had equivalent numbers (through catalog changes). This serves as a basis for determining the unduplicated courses on the transcripts. Certain decision rules must apply in eliminating duplicated courses or in determining how to treat catalog changes. For example, how will you treat cross-listed courses. A course titled "Literature of the American West" appears in the catalog as both a history department course and an English department course. It is part of the interdisciplinary American studies curriculum and is taught alternately by faculty of both departments. Students enroll in these courses either for English or history credits, yet they encounter in any given term, one faculty, one syllabus, one set of readings and exams; in short, one effect. A decision is needed with regard to how to treat this course. In this case, coming from the DCP Project, we determined whether more students from the sample group enrolled for history credit, then treated all enrollees (English and history) as history enrollees. In this way, we created a set of decision rules for analyzing the curriculum and identifying anomalies in the course of study in the process. A set of decision rules is provided as an example in Appendix A.

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

You may have found more than 5,000 courses on the transcripts of 100 students participating in the assessment program. You now have identified and listed only courses appearing on more than one student transcript. You also have decided how you will treat cross-listed courses, catalogue changes, and courses with multiple sections. Now you may still have 1,500 unduplicated courses coming from the students' transcripts. Since the objects of the analysis are those courses that may contribute to general education and liberal learning, you will need to select from that list of 1,500 courses only those who enrolled five students. Why five students? You will want enough students from your assessment group to generalize about the course.

If you plan to pilot test the CCAM, then we suggest that you limit your analysis to courses enrolling 5 or more students. If you are planning a full implementation of the CCAM, then enrollment of 15 students or more per course reduces the probability of error in the model significantly. In all cases, the number of courses you will be able to analyze will be limited by the capacity of your computer and the software you employ. For example, we have found that SPSS-X can analyze no more than 994 courses using the procedures described in this handbook. Normally, this should not be a serious limitation in that the transcripts of 100 students will produce about 5,000 duplicated courses or 1500 unduplicated courses. This then will be reduced to perhaps to 330 unduplicated courses that five or more students took or 150 unduplicated courses that 15 or more students took. In this manner, we move from the universe of all courses students took to those that were taken in common by five or more or 15 or more students. This becomes the common undergraduate curriculum base from which the general education curriculum is embedded. Only those courses most frequently chosen by students should be included in the analysis. Obviously, average class size has a bearing on the number of courses available for analysis, given a specified transcript sample size and minimum number of students required in each course cell in the cluster data matrix.

The Curricular Representativeness of the Coursework Found on Student Transcripts

The correlation between GRE and SAT scores is fairly high (Ratcliff, 1987). The smaller the number of enrolled students in a course the greater the probability of error in calculating mean residual scores. Thus low enrollment courses may distort the information yielded from the cluster analysis. From the results of the DCP Project, we found those courses enrolling at least 5 or more students could be cluster analyzed using correlation coefficients as the metric values.

Unfortunately, in most colleges and universities we don't have a true, clear idea of what the total curriculum is. Courses are added and deleted, and the catalogue does not really represent the courses available to the student in any one term. Under such circumstances it is difficult to generate a sample of the undergraduate curriculum. Certain courses are offered in alternate years; others are offered less frequently. Some courses are cancelled for lack of enrollment; others are split into multiple sections taught by different faculty due to large student demand. The exact number of courses available for enrollment in any given year or term is often not available. The courses in one year are not exactly identical to those offered the following year. What constitutes the curriculum, in terms of number of courses, content and variety, varies from term to term and year to year.

Without an exact definition of the total undergraduate curriculum, the representativeness of a sample of courses can only be approximated. Since the curriculum is changing throughout a student's baccalaureate program, and since undergraduates enter and exit at different terms and times; and since the tenure of their undergraduate studies varies, the exact extent of courses from which a student can make choices becomes individual, nebulous and imprecise.

When the Coursework Cluster Analysis Model has been used with a ten percent sample of graduating seniors at some other institutions, wherein courses enrolling five or more students were examined, the proportion of the total curriculum represented by those courses was significantly smaller than the total unduplicated courses on the transcript. Coursework taken by five or more students was 15 to 33 percent of all the coursework listed on the transcript (Ratcliff, 1988). When the initial sample size is not very large, the representativeness of the courses to the total curriculum may be seriously questioned. However, many debates regarding the vitality of the undergraduate curriculum in producing general learning among students consider only the general education portion of the curriculum, not every course listed in the catalog. From that standpoint, the representativeness of the courses included in the cluster analysis may be defined in terms of either (a) the total of courses offered during the period of enrollment of the student, or (b) the combinations and sequences of courses prescribed by the college or university to meet the general education requirements for a bachelor's degree. That is, when you examine unduplicated courses taken by five or more students, you can determine what proportion of those courses reflect the general education requirements and what proportion of all unduplicated courses on the transcripts are present. The first tells you about the general education curriculum; the second about the total undergraduate curriculum.

The total courses offered during the period of enrollment of a student is not easily ascertained at many colleges and universities. First, the transcripts of a cohort of students list only what the students chose, not what was offered but they didn't choose. Student choice of coursework is not made in isolation, but is made in relation to those courses not selected, those previously selected, and those planned for future terms. Second, not all courses offered during a given period are listed in the college catalog or bulletin. Experimental courses and new courses, some of which may be extended only in one term or year, do not appear in the catalog. Comparing the student transcripts with the college catalog reveals this. Thus, courses not listed in the catalog and not selected by a given cohort of students were among the range of enrollment choices available to the students. Lastly, there are courses in the college catalog which may not be given during the enrollment period of a cohort of students. While such courses were not choices to the student cohort, they were regarded as part of the formal curriculum of the institution. Thus, defining the curriculum as all courses available and/or advertised to a particular cohort of students may not produce an exact representation of the college curriculum. It may, in fact, obscure some of the most experimental and innovative courses which, for one reason or another, did not get recorded in the college catalog.

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

In the end analysis, however, it is net effect of coursework patterns rather than their representativeness that is most important to linking student learning and the undergraduate curriculum. If one wanted a sample of courses representative of all listed in the entire curriculum, one would need a large enough student sample so that the courses appearing on the transcripts of 5 or more sample students would be representative of the total curriculum. This, again, presumes a means for determining the totality of a curriculum, given changes in courses offered over the period of student enrollment. It also requires an investigation of the relationship between the representativeness of courses to the curriculum and the relationship of the representativeness of the sample transcripts to the student cohort or population studied.

At this juncture, it is important to note that the focus of the Model and its accompanying analysis is on courses, not students. It is not the purpose of the cluster analysis to predict the population mean parameter of all the students enrolled in a course. Since the main purpose of the cluster analysis of college curriculum is to examine the effect of an unknown course enrollment pattern on student general learned abilities, the confidence level of mean residuals for an individual course is not of much importance because the attributes are in large part significantly determined by all students in the sample group, rather than by the students enrolled in that course alone. Thus, there is no reason for deleting those courses enrolling 4 or less students from the cluster model building because the course attributes are determined by student course enrollment pattern, not by the characteristic of a single course. The principal reason for restricting analysis to courses taken by five or more students should be the discovery of coursework taken in common rather than individually by students.

Some Limitations in Analysis of Curricular Patterns

By analyzing coursework that leads to higher student gains in general education and liberal learning, you should be able to identify the net effect of different parts of the undergraduate curriculum. That analysis should also point to those parts of the curriculum which promise to be most effective for promoting student learning and cognitive development.

Two types of error need to be avoided in such an analysis. The first is the ***reductionist error***. Here you error by attributing the score variance in complex groupings of coursework to individual psychological variables. For example, you may analyze the differences in learning between African-American and Anglo-American students, noting that students in each group who showed large gains took different coursework. It would be a reductionist error to assume that the coursework and learning differences were solely attributable to race alone (Grant & Sleeter, 1986). The reductionist error may also occur in research equating general learned abilities within complex academic organizations with intra-group cohesion and/or with individuals' identification with an academic discipline. The study of student learning in colleges and universities is a study of student behavior <u>in</u> such organizations, rather than the study ***of*** such organizations.

A second type of error is to presume the ***uniqueness of the data***. You may note differences in learning among African-American and Anglo-America students associated with differences in coursework chosen. However, it would be an error to presume that only these combinations of courses produce these effects. In fact, the analysis identifies the courses associated with the most gains in learning from the universe of course combinations tried, not from the universe of course combinations yet untried by the students. While it is important to acknowledge what is unique in each institutional learning environment, this should not halt the exploration of appropriate alternative relationships of curricula within different colleges and universities. This error may emanate from the failure to conceive of these institutions as systems (1) nested within and linked to larger systems (disciplinary and professional fields), and (2) containing smaller subsystems (departments, divisions and programs) that are, in turn, linked to them (Katz, Kahn & Stacey, 1982).

Prior research suggests that student coursework patterns found to affect general learned abilities can be characterized by (1) the extraneous (other than achievement) characteristics of the students enrolled, (2) the unique or idiographic characteristics of the learning environment, and (3) the normative effect of the fields of study on learning in colleges and universities (Astin, 1970a; Pascarella, 1985). Prior research has also demonstrated that more than one model of college curriculum can explain the effect on student learning from a common set of transcript data (e.g., Hesseldenz & Smith, 1977; Kolb, 1973). Therefore the Coursework Cluster Analysis Model identifies empirically-derived course patterns which subsequently may be examined in terms of student characteristics and idiographic aspects of the curriculum. In this sense, the cluster-analytic model is retro-deductive in approach and is useful to the generation of research questions and hypotheses regarding common notions of the college curriculum and its relationship to general student learning at the undergraduate level.

Summary

This chapter described the conceptual framework used in analyzing coursework patterns. The coursework patterns identified from student transcripts represent the dependent variables in this design. The assessment scores of students (derived from the selected assessment instruments) represent the independent variables. The Coursework Cluster Analysis Model serves as the vehicle through which to link the coursework students take with their improvement in learning. The next chapter presents an overview of some major considerations in planning assessment programs of student learning.

CHAPTER THREE
SOME CONSIDERATIONS IN PLANNING FOR THE ASSESSMENT OF STUDENT LEARNING

It is difficult for the college community to reach a consensus among its constituents on the goals of general education or on the best instrument to use in the assessment of general abilities.

> General education involves many tensions: between what to teach and how to teach it, between the great classics of the past and contemporary works, between the classroom and students' out-of-class life, between students' individual objectives and the needs of the community, between what students want and what their institution think they need ... (AAC, 1988, p. 5).

First and foremost the curriculum serves the students. The quest for an ideal measure of general learning can be revisualized as a search for multiple measures which appear to be appropriate criteria for describing one or more aspects of the general education goals of the college. Once potential instruments have been identified, they can be tested using the Coursework Cluster Analysis Model. CCAM will show the extent to which each measure explains general student learning within a particular college environment.

Gaining Faculty Support and Institutional Commitment

A challenging and yet critical task is getting the support of concerned constituencies. The active and public support of senior executive administrators especially the president and chief academic officer is crucial. However, if assessment is viewed as a priority within the institution then everyone needs to be involved including faculty and students. All of these groups should participate in discussions about the purposes, uses, and benefits of assessment. Administrators can create an atmosphere of trust in which faculty feel comfortable to discover curricular and instructional strengths and weaknesses. If adequate time and attention is given to public meetings about assessment goals and plans, then the fears of faculty and others can be diminished. Faculty often fear that assessment will focus on negative evaluations and believe that student outcomes are unmeasurable (Ewell, 1984). Assessment should be presented as a developmental process and a mechanism for both individual and institutional improvement.

Faculty commitment to assessment will increase if the faculty are involved in the process from the very beginning and in the design and implementation of plans. They especially need to participate in the interpretation of results and development of recommendations. Respected faculty leaders should be involved and faculty members with technical expertise in research design or measurement can serve as consultants. Without faculty support, success of an assessment program is unlikely. Faculty members are a critical source of both political and technical support. Thus, faculty should be recognized and rewarded for their participation.

Setting Objectives for the Assessment Plan

Prior to making a selection of assessment instruments, an effective assessment program begins by establishing concrete and measurable educational objectives. Many college catalogues discuss institutional goals, purposes or missions in the form of broad ambiguous statements, such as developing character or cultural appreciation. Goals are normally abstract and global. However, objectives serve to indicate in concrete terms what the specific program, in this case general education, is attempting to accomplish in the evaluations of student learning. These objectives identify what specific skills students should possess or develop (including higher-order cognitive skills) and what students should know in terms of content. Objectives represent the intended outcomes while the assessment results provide an empirical basis for determining whether these actual outcomes are achieved. Gardiner's *Handbook on Planning for Assessment* (1989) contains practical advice regarding the development of specific objectives. The determination of objectives for general education programs are important and serve to guide the selection of assessment instruments to evaluate student learning. Assessment is a continuous process which systematically compares an institution's performance to its purpose.

Identifying Assessment Instruments

Once a college commits itself to a comprehensive program of assessing general education and liberal learning then potential instruments need to be identified. Institutions can choose from commercially-available standardized tests or develop local instruments. They will also need to choose between norm-referenced or criterion-referenced instruments. In norm-referenced instruments, an individual's score is interpreted by comparing it with other students' scores obtained on the same instrument. Criterion-referenced instruments use a specific content domain as its interpretative frame rather than a specified population of persons. The focus is on what the person can do and what knowledge or skills have been mastered.

Banta and associates (1990) have developed a ***Bibliography of Assessment Instruments*** which provides a brief profile of eight different standardized instruments to assess general education. Information for each instrument includes the publisher, scales, length of test, time to complete, the cost, and additional references pertaining to the specific instrument. Profiles of instruments reviewed include the Academic Profile, the ACT Assessment Program, the College Basic Academic Subjects Examination (College BASE), Collegiate Assessment of Academic Proficiency, CLEP Education Assessment Series, CLEP General Education Examinations, College Outcome Measures Project, and the General Examination of the Graduate Record Examination. Banta and associates also publish a newsletter, ***Assessment Update***, which contains articles that frequently critique available instruments and discuss relevant assessment issues.

A specific example of a standardized examination for selected college level content learning is the College Basic Subjects Examination (College BASE). College BASE is a criterion-referenced achievement test "intended to assess content knowledge and skill development commensurate with student completing the general education component of their college experience" (Osterlind, 1989, p. 1). The test provides a composite score and four content subscores: English, mathematics, natural sciences, and social studies. The College BASE is currently being used with the Coursework Cluster Analysis Model on an experimental basis at the University of Tennessee, Knoxville. The Model has also been used with the Kolb Learning Style Inventory.

Standardized instruments are useful in assessments since the publishers have usually established some level of technical confidence and furnish norms that permit score comparisons. They are particularly advantageous in the assessment of a large number of students. However, these instruments have disadvantages, one of the largest being that they may not examine what faculty are teaching. They may assess primarily lower-order intellectual skills. Faculty may fear that the assessment may influence what is taught when the emphasis of curriculum reform is directed by the assessment results. These instruments may be standardized on norm groups which are not representative of the population, and they often provide few subscores which makes it difficult to decide what to change when scores are lower than expected (Banta, 1990).

Faculty may develop assessments of student learning in areas they teach. Locally-developed assessments of general education may be more difficult to compose because they often involve several disciplines, fields, or general concepts of cognitive development, such as critical thinking, rather than specific learning in the disciplines with which faculty are readily familiar. Nevertheless, because such instruments or measures are designed by those who did the teaching, these instruments are more likely to test what is actually taught by faculty. A disadvantage is the amount of time necessary to develop the instrument. In constructing such instruments faculty gain additional expertise in developing evaluations. This can enhance the quality of their own assessment instruments used in their individual courses.

Whatever instruments or measures are chosen, they should match the purposes of an institution's general education program. Different assessment instruments of education reflect different conceptions of general education. Similarly, general education can embody a wide array of educational objectives which differ greatly depending on the institution's mission. An institution can select any instrument which measures its particular view of general education.

Selecting Instruments: Some Technical Guidelines

The reliability and validity of assessments is critical to the success of the assessment program. Reliability refers to consistency and stability of scores or results obtained by the same persons when re-evaluated with the same assessment instrument on different occasions, or with different sets of equivalent items. If an individual wanted to assess a particular student's development, then a number of observations would need to be collected or a number of test questions would need to be posed to achieve a greater degree of dependability. The process of gathering several observations or measures about similar sets of educational objectives strengthens the degree of reliability. Likewise, validity is critical. It refers to the accuracy and appropriateness of the assessment instrument results and their interpretations and use. How well does the instrument measure what it is supposed to measure? Here are some additional important things to consider in choosing assessment instruments (Erwin 1990, p. 76):

- "Is there evidence for reliability? Is the type of reliability appropriate for the type of assessment method?
- Are the items of the test or rating scale keyed to specific program objectives? Are the items keyed to the developmental levels, such as application or analysis?
- Have gender, cultural, ethnic, and geographical biases been minimized?

- Are the directions for administration clear? Are the rules for scoring the responses and behaviors clear?

- Is the sample for standardization and initial design of the instrument clearly described?

- Are percentiles provided, with sample descriptions, for norm referenced interpretations?

- Are mastery cutoff scores or rating levels provided for criterion-referenced interpretations?

- What evidence for validity exists? Is the level of difficulty appropriate for the intended program?

- Are the scores or ratings reported in a form that is useful for the program? For example, some tests report only a total score, which many not be useful for diagnostic purposes. What is needed are subscores for assessing various components of the program."

Answers to these questions can usually be obtained from information included in the publisher's assessment materials.

Another method for determining the appropriateness of particular assessment instruments is to have experienced faculty conduct a content analysis of the potential instruments prior to a pilot study (Banta & Pike, 1988). The results from this activity suggests the preliminary degree of correspondence between the content (items) of the assessment instrument and the statement of desired outcomes of an institution's general education program. Several faculty members can individually conduct this analysis and then in a group discussion they can work towards a consensus of which instruments appear to measure the outcomes of the general education program.

In thinking through what instruments or measures to use in the assessment program, be liberal and expansive at this point. Encourage faculty involved in the instrument selection and/or development process to think creatively and not be too concerned with overlap and duplication of measures. Later the Coursework Cluster Analysis Model will tell you which measures best describe general education and liberal learning at your college. Also, a pilot test of the instruments with students will provide valuable insights in how well they will provide the needed portrait of student learning.

The fundamental questions that should guide your process of selection and/or development of instruments are: (1) do the instruments measure what we are trying to achieve in general education and, (2) will they consistently report what we want to know. Remember that instrument development is a time consuming and expensive process. We suggest that you begin with a review of available commercial instruments. Next, you can fill in the gaps. That is, as you review your general education goals you will probably find some currently available instruments which provide some of the assessment information needed to give a picture of student learning relative to the general education goals. However, these instruments probably will not give a total picture. Here is where local instrument development should begin. You don't need to put the entire assessment program on hold while these additional measures are being developed. The CCAM allows you to add additional measures as you proceed through the assessment process. As long as you begin the process with faculty understanding that the current set of instruments or measures is a *partial* portrait of general education and that additional or revised measures will enlarge the picture as time and experience accumulate.

Conducting Pilot Studies

The Coursework Cluster Analysis Model provides information to guide the choice of assessment measures. Remember multiple measures or instruments were selected with the understanding that no one particular instrument is perfect for the job. Using the Model and accompanying analyses, you can determine the extent to which the selected measures may overlap in describing student learning and which mix of assessment measures are most appropriate given the general education goals of your institution.

Such a small pilot study will help organize and improve the larger more rigorously defined future assessment program. It will help you determine which instruments are most appropriate for the assessment program. A pilot study will also inform and help refine procedures used in the assessment program. Interviews with students who have completed these assessment instruments may also be useful to determine their perceptions of the items correspondence with the outcomes of general education. A sense of the students' motivation to perform their best effort on the instruments can also be ascertained. Through conducting a pilot study using different multiple instruments, the future larger scale assessment effort will be strengthened. Institutions will obtain empirical information concerning the extent of variation in student assessment results that is explained by the measures used in the Coursework Cluster Analysis Model.

A serious issue in any assessment program is the students' motivation to give the best effort to complete the assessment to the best of his or her ability. If there is no reward for students doing their best, the validity of the assessment information is diminished.

There are two basic ways to get students involved in the assessment program. You can reward them and/or you can require them to participate. Students planning to attend graduate or professional schools may be rewarded by taking standardized tests used in graduate admissions such as the GRE, Miller's Analogies, MCAT, LSAT. While students planning to continue their studies may be eager to do well on these tests, others may not be eager. The objective of the assessment program should be to get an accurate picture of all students at your institution rather than just those planning graduate studies.

Students have been motivated to participate in the assessment program by an invitation from the college president. Here the president appeals to the students' sense of loyalty and institutional commitment. Also, the president may promise to write letters of recommendation to future employers for those who do really well on the assessment measures (Morante, 1991).

Financial incentives induce students to participate in an assessment activity who otherwise would not do so. Some projects (including the Differential Coursework Patterns Project) have paid students a stipend that is slightly over minimum wage or what they might earn on-campus at work study jobs. Some have used rebates or coupons for food or college services in lieu of cash payments. This avoids complicating the needs analysis for students on financial aid. Some assessments have successfully used lotteries, wherein all who participate are entered into a drawing for prizes such as a spring break expenses paid vacation or dinner with the college president. Financial incentives encourage students to participate who would see less direct academic benefit or reward for the assessment.

When the assessments are administered and how long they take also has direct bearing on the success of the program. A college serving a commuter student population may have real difficulty convincing students to return to campus on a Saturday for a half day or an all day assessment program. Assessments held during the regular instructional week that are at times and durations equally convenient to traditional-age day students and adult and part-time evening students are more likely to succeed.

Pilot testing will tell you much about the instruments you selected, the procedures you used, and the extent to which the students who participated really represented the college student population as a whole.

Considering the Costs Involved

The more you want the assessment program to tell you, the more it will cost. Setting modest goals for one program at the outset may be the best approach. Given the flexibility of the CCAM, the program can be expanded in scope as new applications or new information is sought. The extensiveness of the program (including the size of student samples) will affect the amount of money which an institution will need to invest. The major costs for an assessment program using the Coursework Cluster Analysis Model will be in four major areas. First, the instruments selected to assess student learning will vary in cost. Locally-designed instruments may cost more in the short-term due to the time needed to develop these measures. Second, there will be administrative costs for the assessment including student motivation to participate costs. The third area will be costs associated with conducting the analysis (including personnel and computers). Fourth there are costs associated with the overall coordination of the assessment program. Institutions which are just beginning their assessment programs will also incur more costs initially. Special resources such as start-up monies will be needed for key staff to attend assessment institutes, to visit other institutions, to hire consultants, or release faculty time to develop instruments and participate in overseeing the assessment process.

In any significant college or university activity, devoting five percent of the costs to determining and evaluating how beneficial and effective the activity is reasonable. Few colleges devote as much as five percent of the instructional budget to determining effectiveness of the instructional program. Yet such a commitment is needed if we are to take teaching and learning serious. A fraction of that five percent should be allocated to an assessment of general education. The institutional costs in terms of both time and money associated with assessment represent an investment in the future of the institution and its students. There are numerous benefits in conducting an assessment of a general education program's outcomes. The process will help to clarify the educational program's goals; the curriculum may be revised to improve student learning or confirm the value of its current structures; academic advising can be enhanced; and ultimately faculty may be motivated to revise their own individual courses or assessment techniques based upon the results gained from an ongoing assessment process.

Issues in Selecting Student Samples

The assessment process is more than collecting data from an examination given to volunteering undergraduates. Before the data can be gathered, major decisions must be made about who will be assessed, how often, and when these assessments will occur. A major goal is usually to assess a subgroup or cohort of students from a college with the expectation that the results can be generalizable to the student population at the particular institution. Samples of students can be studied over time and assessment information may be collected several times such as during the freshmen and seniors years.

Frequently it is too expensive to assess every student particularly at large institutions. Therefore, academic leaders may want to identify and select at random a group of students who would represent the characteristics of all the students in a given institutional environment. The demographic characteristics of a random sample of college students should be compared with the same demographics of the entire student population at the particular institution. For example, in the work conducted using the CCAM, a thorough comparison of the student sample was made to the population by examining students' major, gender, ethnicity, grade point average, and age. This systematic comparison indicated that the sample's characteristics were analogous to the population of graduating seniors (Ratcliff, 1988). This is not the same as the population of students enrolling at the institution. Seniors represent only those students who will likely finish their degree program. A different sampling strategy would be needed to incorporate those who drop out or did not complete their degree program. The selection of student samples should be carefully made to insure that they represent a given population of students.

Summary

This chapter highlights major areas where academic leaders must make decisions which ultimately impact on the design of the assessment program for general education. It is critical that institutions have concrete and measurable objectives for assessment. These objectives determine and directly influence the selection of measures or instruments to assess student learning. Pilot studies are worth an investment of time and effort in order to assess the results of a variety of measures on a trial basis. An important issue is student motivation to complete to their best ability the assessment. The participation of students, especially in the assessment planning process, can strengthen student motivation and generate realistic expectations for assessment. The costs associated with the assessment program will vary considerably depending upon the scope and size of student samples. Through careful and thorough planning in these major areas, an assessment program of general education can subsequently be implemented, revised, and expanded as the academic leaders continuously review the assessment results. The next chapter describes how to determine the relationship between coursework patterns and assessment scores in order to identify the areas where students gain in their learning.

54

CHAPTER FOUR
METHODOLOGY AND PROCEDURES: COURSEWORK CLUSTER ANALYSIS MODEL

Cluster Analysis and Student Learning

The Coursework Cluster Analysis Model is grounded conceptually to the finding that student learning varies more greatly within institutions than between them. The selection, testing and adoption of a specific methodology for the analysis of coursework patterns was based also on repeated empirical investigation of the relationship between different patterns of coursework and variation in student learning. In this chapter we describe the general methodology of the CCAM. The rationale and procedures of cluster analysis are described with reference to its application to the investigation of coursework patterns. We contrast cluster analysis to other statistical methods of potential value in the assessment of student learning. Chapters Five and Six describe in detail a step by step implementation of these CCAM quantitative procedures.

Definitions of Analytic Techniques

Previous assessment and transcript analysis studies have used the general linear model and *regression analysis* (Austin, 1970a, 1970b; Benbow & Stanley, 1980, 1982; Pallas & Alexander, 1983; Prather & Smith, 1976a, 1976b). The rationale for the use of regression is based upon practical and theoretical justifications. Regression analysis allows maximum design flexibility and is statistically robust. Transcript analyses involve large amounts of data. For example, Prather et al. (1976) examined 8,735 student transcripts which collectively contained 189,013 individual course grades. Regression analysis provides an effective technique for presenting the diverse nature of the data while maintaining a consistent analysis rationale. However, the general linear model does not provide a direct means of assessing the additive and temporal aspects of course patterns, as described in the previous chapter. Furthermore, use of linear regression alone would conceptualize the problem as finding the one best fit between students and learning experiences. It would not account for different learning experiences being appropriate and beneficial for different groups of students.

What do we mean by coursework? We used this term to refer to the categorization of the courses in which students enrolled according to the multiple assessment criteria of their general education and liberal learning. It is the systematic and unique way a college or university labels and arranges its courses (i.e., Honors 101, French 340, etc.); that scheme or arrangement of classes is already known in a disaggregate form on student transcripts. Identification is the allocation of individual courses to be established in categories on the basis of specific criteria (i.e., Biology 205 is classified by many universities as a sophomore level class in the department of Biology).

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

Cluster analysis is sometimes confused with factor analysis. Factor analysis is different from cluster analysis in that its attention is on the similarity of the variables (attributes). The aim is to identify a small number of dimensions (factors) that can account for individual differences on the various measures or attributes. Thus, the aim of factor analysis is to reduce or consolidate the number of attributes of a variable set while the purpose of a cluster analysis is simply to classify or taxonomize data into groups on the basis of a set of attributes. Miller (1969) examined 48 common nouns; through cluster analysis he identified five subgroups referring to living things, non-living things, quantitative terms, social interactions, and emotions. Another example of cluster analysis is Paykel's (1971) analysis of 165 depressed patients. Using symptom ratings and historical variables, he grouped the patients into four clusters: the retard psychotic, the anxious, the hostile, and the young depressive. Cluster analysis refers to a wide variety of techniques used to classify entities into homogenous subgroups on the basis of their similarities.

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

Cluster analysis conforms to the conceptual restrictions placed on the CCAM to assess the effect of coursework patterns on student learning. Cluster analysis provides a statistical procedure for examining coursework using multiple criteria. It can classify different sets of coursework according to different net effects of learning associated with them. It can accommodate both quantitative and qualitative attributes of varying dimensions. Thus, the criterion selected need not be test scores; nominal, order, interval and ratio data have been successfully used as attributes in cluster analysis (Romesburg, 1984). Cluster analysis uses these attributes to arrive at patterns of coursework independent of any institutionally prescribed ***a priori*** distinctions. Therefore it can test the combinations, sequences and progressions of courses within the undergraduate curriculum. It leads to the discovery of clusters (or patterns) of coursework in student transcripts, based on the multiple measures of student assessment employed. Since the purpose of the CCAM is to group coursework homogeneously relative to student learning criteria (Lorr, 1983; Romesburg, 1984), cluster analysis serves as the primary methodology for the analytic model.

CCAM Procedural Steps

There are several steps to using the CCAM. First, student residual scores are derived. Next student transcripts are examined. Courses reported on them are clustered into patterns based on the residual scores of the students who enrolled. The resulting coursework patterns are then grouped or classified according to any of a wide variety of student or institutional factors. Patterns can be classified according to factors such as the entering ability level of the student, the type of coursework selected (general education, major, minor, prerequisites), the campus at which the student enrolled, or the residence facilities housing the students. Adult versus traditional college age students; commuter versus residential students; and part-time versus full-time students' coursework can be compared. Within systems of higher education with course comparability, transfer schemes, and articulation agreements, CCAM can be used to determine if coursework associated with students from branch campuses or transfer students are associated with the same types of improvement in learning as for students native to the campus.

Hypothesized patterns of coursework generated from one set of student transcripts may be validated through the replication of the Cluster Analytic Model to a second sample of student transcripts.

Deciding How Much Curriculum to Monitor: Using Quantitative and Qualitative Measures

There are at least two views of what constitutes representation of a college or university curriculum. One view holds that only those courses in which students most frequently enroll constitutes the curriculum associated with general learning. A second view posits that any course offered may contribute to the general learning of students. The first view implies a more restricted view of the curriculum than does the second. Eachview requires different CCAM procedures. We describe procedures to determine the effect associated with coursework in which students most frequently enroll. The multiple assessment measures selected for use with the CCAM serve as attributes for the classification of courses into patterns. These attributes can be expressed quantitatively or qualitatively. The mean of residual scores for students enrolling in a sophomore level mathematics class (for example, Math 201) can be described according to the mean score gain of students (from the sample) who enrolled in the course. The course mean residual is a quantitative attribute. Math 201 can also be described nominally; here the researcher simply notes whether one or more students with high score gains enrolled in the course. Both the quantitative and qualitative descriptions of Math 201 serve to determine the relation of the course to other courses according to the assessment criteria used.

When a *sample* of students is used for the analysis, not all courses in the curriculum appear on the transcripts of the student sample. When the sample is small there are a limited number of courses within the curriculum which can be analyzed quantitatively. Also, only a limited number of the courses appearing on the sample transcripts can be analyzed if the number of students enrolling in a given course is a concern in the analysis. For example, if you want at least 10 students from the sample in a course before you include it in the analysis, you will eliminate a number of small class size courses from the analysis, particularly if you are using a sample of students. Using a class size of 15 reduces chance error in the CCAM further. But the precision of information generated from such analysis may be compromised by the limited proportion of the curriculum that is reviewed. Quantitative analysis of the curriculum can yield much more accurate information regarding the effect a particular course may have on a given measure of student general learned ability. To generalize about a course on the basis of 5 or more student score gains provides a level of information that far exceeds that of simply noting whether any student who performed well on a given measure enrolled in that course.

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

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

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

In a quantitative cluster analysis, the metrics used for each course are the mean score residuals. Course mean residuals contain interval information about improvement in student learning for those who enrolled in the course. In a qualitative cluster analysis, the metrics used are whether students with high score residuals did or did not enroll; the metric is reduced to a dichotomous nominal variable. There is a trade-off in a qualitative cluster analysis between inclusiveness of the curriculum and precision of the information.

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

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

Second, institutional researchers have a choice between an intensively detailed picture of the curriculum using the ratio data of mean residuals or a less detailed picture provided by binary information. If the primary goal is ***individual student assessment***, then an appraisal of the learning of all students is warranted. Assessment of an institutional curriculum or program variables may only require a sample of students to generate the information needed for such analyses. There are occasions when the scope of the analysis is to be preferred over the precision of the analysis.

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

Cluster Analysis using Quantitative Measures

Described below are the steps required in the CCAM quantitative analysis to assess the effects associated with the coursework patterns on the general learned abilities. The research design uses as data sources transcripts and instrument scores from a sample of students. The nine item-type categories of the General Tests of the Graduate Record Examination are used as measures of general learned abilities of college seniors for our example. Again, standardized and non-standardized, locally-developed and commercially available assessment instruments and measures may be used with CCAM. In our example, SAT scores are used as controls for the academic abilities of these students when they first entered college. The student transcripts are used as the record of the sequence of courses in which these seniors enrolled.

The first objective of the CCAM is to determine the extent of student improvement in general learned abilities over the time of their baccalaureate program. To do this, first the residual score of each GRE 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. It is derived by regressing the outcome measure (in this case GRE item-types) on the entrance measures (in this case, SAT scores). Thus, for each student outcome measure there is a student residual score for each person in the sample group.

The second objective is to determine patterns of coursework on the student transcripts which are associated with the student score residuals. Cluster analysis gives us these patterns, using student residual scores (GRE item-type residuals) as attributes of the courses in which students enrolled. To do this we create a data matrix where all the courses to be analyzed are in columns and all the assessment measures or criteria are in the rows. Each cell in this matrix is then filled with the appropriate mean course residual score. For example, let us assume that we have student assessment data on student writing ability, understanding of scientific knowledge, and a writing sample that has been holistically scored. For the course Introduction to Political Systems, we calculate the mane of residual scores for all students enrolling in it for each of these measures. We do this for Introduction to Political Science and every other course on the students' transcripts that we select to analyze.

Now, with several rows of assessment data, and a column for each course analyzed, and a course mean residual score in every cell of the data matrix, we are ready to determine how similarly students who enrolled in different courses performed. The course mean residual score is the metric value we are going to use to make the comparisons of coursework. To determine how courses are similar to one another in this way, we use the correlation coefficient (Pearson's *r*) as the indicator of similarity.

Our task is to see how the performance of students in our course Introduction to Political Systems is similar to the performance of students in other courses. However, students take more than one course, so courses that a group of students who showed large improvement took will group together. That's because the course mean residuals for each assessment measure should look about the same for all the courses a group of students took.

So if we correlate the writing sample score of Introduction to Political Systems with the sociology course Mass Behavior, then the correlation will be high if students for both courses showed comparable improvement on that measure. What we are doing, then is creating a second matrix to record our correlation coefficients. In this matrix, all the rows are the courses analyzed and all the columns are a duplicate listing of all the courses. Each cell contains the coefficient representing the extent to which each course is related to all other courses or all the assessment criteria. Obviously, the greater the assessment criteria the more precision in establishing the relationship. These two data matrices, the raw data matrix and the course resemblance matrix, sound like a lot of work. Fortunately, the computer using popular statistical programs, such as SPSS and SAS, will do this for you. You won't even see these matrices as they are calculated in lightning speed as you move along performing the CCAM cluster analysis.

Once the ***resemblance matrix*** indicating the proportional relationship of courses is established, a clustering method is selected and executed to arrange a tree or dendrogram of courses related by the student score gains. Next, we conduct a discriminant analysis on the resulting clusters of coursework. The discriminant analysis tells us (a) the extent to which the courses have been correctly classified according to the assessment criteria, (b) which of the assessment criteria were correlated with particular discriminant functions, and (c) which coursework clusters were associated with the improvement of student learning according to which assessment criteria. From the discriminant analysis an association can be inferred between coursework patterns (clusters) and the assessment criteria (student score residuals on the multiple measures of learning). The cluster-analytic procedure groups courses frequently chosen by students according to the strength of their associated effect on the student score gains.

Summary

The CCAM classifies the most frequently enrolled courses according to their associated effect on student improvement in learning. The quantitative procedure classifies courses according to a ratio index of similarity to other courses. Procedure is designed to examine those courses in which most students enroll. Thus, the analysis is limited to only a fraction of all the courses in a college curriculum. For example, in the historical database used in model-building and testing, a five percent sample of student transcripts enabled an examination of only five percent of courses appearing on those transcripts (the percentage of courses enrolling 5 or more students from the sample group). However, the courses examined in that 5 percent corresponded closely to those courses identified as meeting the College's distributional degree requirements in general education. The remaining chapters in this handbook provide a detailed presentation of how to conduct the quantitative procedures.

CHAPTER FIVE

DETERMINING WHAT STUDENTS HAVE LEARNED

Any assessment is only as good as the measures and indicators used to make the assessment. In this chapter, we describe how to use the CCAM to determine the reliability of the assessment measures selected. We describe how to use the CCAM to determine if the various measures used in the assessment are measuring similar or different types of student learning, and how to calculate student residual scores as proxies of the improvement in student learning. Our specific examples use student SAT scores as the measures of their entering abilities and the GRE nine item-types as measures of student general learned abilities as they complete their baccalaureate program. However, please remember that various assessment measures can be used to monitor entering student abilities or student learning outcomes. They can include a mix of quantitative and qualitative measures as well as commercially developed and locally-developed instruments. Some of the analyses presented in this chapter may not be necessary depending on the design of the assessment program for general education. For example, if the relationship between the measures of incoming student abilities and the measures of exiting student abilities are known, it would not be necessary to conduct the correlations analysis as described in this chapter. Uses of certain measures may require supplemental analyses as well. For example, if a holistically graded writing sample that is scored on a scale from 1 (excellent) to 10 (failure) is mixed with a multiple choice test graded on a scale from 1 to 100, the scores across measures would have to be standardized prior to making comparisons, otherwise the multiple choice test would carry 10 times the weight of the writing sample in the ensuing analysis.

Reliability of Measures

A fundamental question regarding assessment measures is whether they reliably measure the learning they purport to measure. For locally-developed instruments, this means, "Do the instruments *consistently* describe a kind of learning that this college or university intends to impart to students?" Standardized or commercially-produced tests may be reliable for other student populations but not for the student population at your college or university. Thus, it is important to first determine the reliability of the instruments and their sub-tests for the sample group. Three factors typically contribute to the reliability or unreliability of test scores (Ebel, 1972). The first factor is the appropriateness and definitiveness of the questions. On one hand, the appropriateness of the questions is presumed by the widespread acceptance of many standardized instruments. On the other hand, the appropriateness of the items and item-types or subparts may not measure the type of learning described in the college's goals for the general education curriculum. In this sense, the reliability of a particular standardized instrument may vary from institution to institution.

A second factor contributing to the reliability of assessment scores is the consistency and objectivity of the person (or in some cases, machine) who scores the examinations. Frequently test responses from standardized instruments are read by an optimal scanner and scored by a computer at major testing centers. Consequently, issues of consistency and reliability tend to be mechanical in nature. "What is the error rate of answer sheets read by your institution's (or the testing company's) optical scanner?" Or, "does your error rate go up when answer sheets have been folded or when they have been completed using an inappropriately leaded pencil?" This questions may seem pretty mundane, but they do impact the reliability of commercially-produced assessment scores. On the other hand, a locally-developed and administered student essay or writing sample may present different reliability problems. How do you know that the reader and evaluator of that writing sample will apply equally vigorous and uniform standards to the assessment of the first and last essay read? This can be done by using multiple readers of the writing samples and then examining the inter-reader reliability. Be it quantitative or qualitative assessment informant, the reliability of the evaluator or evaluation process constitute a second reliability factor to designing an assessment plan.

A third factor contributing to the reliability is the constancy or stability of a student's ability to perform the tasks presented in the test. Students may vary from hour to hour or from day to day in their alertness, energy and recall; these may affect test performance, reducing the reliability of the scores. Commercially available assessment instruments come with student groups in order to provide steps to insure uniformity in the information gathered. If your college or university is considering the development of locally-produced measures of student learning, the development and testing of a procedures manual for the administration of instruments and the collection of data should be produced. Remember, the quality of the information generated from the assessment program is only as good as the information gathering process.

Reliability is not merely the property of the instrument or measure itself but also of the measure's individual item-types or subparts of the measure as well. The more appropriate the measure is to the group of students, the higher the reliability of the scores. Ideally, the reliability of a set of scores for one assessment measure may be determined using the correlation coefficient between that set of scores and another set from an equivalent test of the members of the same group.

If the same group of students produces comparable results on two or more administrations of the same measures (or similar forms of that measure), then the test-retest reliability of that measures has been established for that student group. However, it is rarely feasible or practical to give students the same or similar measures twice. Consequently, an alternate means of determining reliability has been to split the student group in half (for the purposes of analysis only). If the scores of the first random half are comparable to the second random half, then the measure or instrument is probably gathering consistent information across the group as a whole. This is the Guttman Split-Half method which estimates reliability by splitting the sample into halves and determining the correlation between the scores in the two groups. The results of the split-half method are dependent upon the manner in which the group is halved. Cronbach's alpha is a statistic designed to overcome this problem. It is a generalized formula representing the average correlation obtained from all possible split-half reliability estimates.

A determination needs to be made about the level required for the reliability coefficients to be satisfactory. For the purposes of the project when conducted by Ratcliff and associates (1988), reliability coefficients at or above a = .65 were deemed satisfactory (Mehrens & Lehmann, 1969). Due to the exploratory nature of this research, lower reliability coefficients were accepted. An example of results for the reliability analysis for the Western sample is presented in Table 1. In this sample, Logical Reasoning (a = .51) evidenced low reliability. The reliability of the individual item-types tended to increase with the number of items comprising the given item-type. Therefore, Analytical Reasoning tended to have a higher reliability since this area contained 38 questions while Logical Reasoning tended to have a lower reliability since it consisted of 12 questions.

Table 1. Sample Reliability Coefficients for GRE Item-Types

GRE Item-types	Code	Number of items	Cronbach's Alpha	Guttman's Split-half
Analogy	ANA	18	.6491	.7563
Sentence Completion	SC	14	.6662	.4118
Reading Comprehension	RD	22	.7090	.7821
Antonyms	ANT	22	.8504	.7941
Quantitative Comparison	QC	30	.8373	.8598
Regular Mathematics	RM	20	.7806	.7360
Data Interpretation	DI	10	.6829	.6155
Analytical Reasoning	ARE	38	.7958	.6535
Logical Reasoning	LR	12	.5074	.5577
GRE Verbal	GRE-V	76	.9125	.8919
GRE Quantitative	GRE-Q	60	.9040	.8964
GRE Analytic	GRE-A	50	.8018	.7299

Correlation of Measures

It is important to determine the extent to which two assessment instruments are correlated; in this case how the GRE item-types and SAT sub-scores are correlated. For example, determining whether the GRE item-type, Analogies, has a stronger correlation with SAT Verbal, SAT Math or the total SAT scores will help determine which SAT score should be used in the subsequent regression analysis. The correlations revealed that the four verbal item-types (Analogies, Antonyms, Reading Comprehension, and Sentence Completion) were strongly related with the SAT verbal scores. The three quantitative item-types (Data Interpretation, Quantitative Comparisons, and Regular Mathematics) were strongly correlated with the SAT quantitative scores. The two analytic item-types (Analytic Reasoning and Logical Reasoning) were strongly correlated with the SAT total scores. Table 2 demonstrates the correlations of the GRE item-types with the SAT scores for the Western sample example.

Table 2. Sample Correlations of GRE Item-Types & SAT Scores

GRE Item-types	Code	SAT Verbal	SAT Math	SAT Total
Analogy	ANA	.5669 ^{4}	.2051	.4469 ^{4}
Sentence Completion	SC	.5503 ^{4}	.2037	.4367 ^{4}
Reading Comprehension	RD	.6267 ^{4}	.4061 ^{3}	.6064 ^{4}
Antonyms	ANT	.5281 ^{4}	.2339 ^{1}	.4431 ^{4}
Quantitative Comparison	QC	.1343	.6024 ^{4}	.4528 ^{4}
Regular Mathematics	RM	.2350 ^{1}	.5616 ^{4}	.4839 ^{4}
Data Interpretation	DI	.3006 ^{2}	.3715 ^{3}	.4016 ^{3}
Analytical Reasoning	ARE	.3834 ^{3}	.4963 ^{4}	.5263 ^{4}
Logical Reasoning	LR	.4348 ^{4}	.3248 ^{2}	.4477 ^{4}
GRE Verbal	GRE-V	.6760 ^{4}	.3210 ^{2}	.5808 ^{4}
GRE Quantitative	GRE-Q	.2353 ^{1}	.6395 ^{4}	.5328 ^{4}
GRE Analytic	GRE-A	.4575 ^{4}	.5154 ^{4}	.5799 ^{4}
Minimum				
Maximum				
Mean				

$^{1}p < .05$ $^{3}p < .001$
$^{2}p < .01$ $^{4}p < .0001$

Intercorrelation of Item-types

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

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

In the example of Western Sample, intercorrelations for Verbal item-types ranged from r = .54 (ANT/ANA) to r = .65 (ANT/SC). Intercorrelations for Quantitative item-types ranged from r = .47 (RM/DI) to r = .60 (RM/QC). Intercorrelations between Analytic item-types were r = .40 (ARE/LR). However, Analytic Reasoning correlated strongly with Quantitative item-types ranging from .51 (QC) to .56 (RM). The intercorrelational analyses showed that in most instances, less than 50 percent of the variance in one item-type was explained by that of another. This finding was relatively consistent with the results from our other samples.

Table 3. Intercorrelation of GRE Item-Types for Western Sample

GRE Item-Types		Code	ANA	SC	RD	ANT	QC	RM	DI	ARE	LR
Analogies		ANA	1.0000								
Sentence Completion		SC	.5717 [4]	1.0000							
Reading Comprehension		RD	.5698 [4]	.6125 [4]	1.0000						
Antonyms	AN	ANT	.5407 [4]	.6474 [4]	.6331 [4]	1.0000					
Quantitative Comparisons		QC	.2351 [1]	.2707 [1]	.3908 [3]	.1372	1.0000				
Regular Mathematics		RM	.2234	.2653 [1]	.4081 [3]	.2128	.6048 [4]	1.0000			
Data Interpretation		DI	.2464 [1]	.4350 [4]	.5175 [4]	.2573 [1]	.5228 [4]	.4714 [4]	1.0000		
Analytic Reasoning		ARE	.2557 [1]	.2675 [1]	.4923 [4]	.2903 [1]	.5111 [4]	.5580 [4]	.5575 [4]	1.0000	
Logical Reasoning		LR	.3998 [3]	.5329 [4]	.6831 [4]	.4323 [4]	.4297 [4]	.2987 [2]	.5575 [4]	.3960 [3]	1.0000

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

Regression Analysis to Determine Residual Scores

The calculation of a student residual score for each attribute (item-type) helps to control for the student's academic abilities prior to entering college. For example, each of the 4 GRE Verbal item-type scores were regressed on the SAT Verbal scores. Each of the 3 GRE quantitative item-type scores were regressed on the SAT mathematics scores. Each of the 2 GRE analytical item-type scores were regressed on the SAT total scores. These GRE item-type residual scores were referred to as student residual scores, that is, the improvement students showed in general learned abilities from the time they entered college to the time of GRE testing during their senior year.

In the cluster analytic model, the SAT sub-scores (Verbal and Mathematics) and SAT total scores were used as measures of entering student ability. To control for the effects of the incoming ability of students, the predictive effect of SAT scores were partialled from GRE item-type scores. For this, 9 GRE item-type residual scores were developed as follows:

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

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

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

While GRE raw scores were generally and consistently high among the students in our samples, differences among scores appeared when the effect of the precollege learning (as measured by the SAT) was removed. When the theoretical scores (as predicted by corresponding SAT scores) were compared with the students' actual responses (Table 4), students showed the largest improvement on certain item-types (in this example, on Data

Interpretation) and the lowest amount of improved performance on other item-types (in this case, Quantitative Comparisons).

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

Table 4 demonstrates an example of the results from a regression analysis. From one-tenth (Data Interpretation) to two-fifths (Reading Comprehension) of GRE item-type score variation among the Western sample was explained by their SAT scores. All regression functions were statistically significant at .0001 with the exception of Data Interpretation which was significant at .001. In a rare case, a function may not be significant even at the .05 level. When this occurs, the item-type is not a strong measure of student learning. Also, the range of residual scores did vary considerably across GRE item-types.

Using the student residuals obtained from the regression analysis above, the mean residuals for each course enrolling 5 or more students were calculated for all the 9 GRE item-types. Such a procedure does not assume that the specific gains of the students enrolled in each course were directly caused by that course. Rather, the residuals of each student are attributed to all the courses in which they enrolled, and the mean residuals for each course serve 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).

Table 4. Sample Summary of Regression Analysis of GRE Scores

Dependent Variables GRE Item-types on SAT Sub-scores	Code	76 Students F Value	Standard Deviation	Adjusted R-Squared
Analogies	ANA	35.046	2.6347	.3122
Sentence Completion	SC	32.148	2.8084	.2934
Reading Comprehension	RD	47.848	3.7173	.3845
Antonyms	ANT	28.616	4.3191	.2691
Quantitative Comparisons	QC	42.137	5.3342	.3542
Regular Mathematics	RM	34.089	3.3784	.3061
Data Interpretation	DI	11.847	2.2581	.1264
Analytic Reasoning	ARE	28.346	5.8506	.2672
Logical Reasoning	LR	18.551	2.2838	.1896
Verbal (raw)		62.267	11.3160	.4496
Quantitative (raw)		51.195	9.3033	.4009
Analytical (raw)		37.490	7.0729	.3273

$p > F = .0001$

Summary

This chapter discussed how to determine the reliability of the measures selected which were determined from the assessment instruments administered to college students. Information regarding how to determine correlations was also presented. The regression analysis described in this chapter indicates the variance in the residuals or the attributes. The variance in these residuals plays a major role in the cluster analysis discussed in Chapter Six. Those attributes (in our example the GRE item-types) will have a more significant effect in sorting courses into clusters.

CHAPTER SIX
DETAILED PROCEDURES FOR CONDUCTING THE COURSEWORK PATTERNS ANALYSIS

This chapter describes the use of the cluster analytic procedure to analyze coursework and scores derived from the selected instruments to assess student learning. In the previous examples discussed in Chapter Five, the GRE was the post-college assessment instrument and the nine item-types scores were the measures of general learned abilities of students. The regression analysis described in Chapter Five resulted in determining a student residual score for each attribute (item-type).

The first section in this chapter describes the computation of mean residual scores for courses enrolling a certain number of students ascertained from student transcripts. In previous work using the Coursework Cluster Analysis Model, the criteria was five or more students from the sample group enrolled for each course (Ratcliff, 1988). This criteria was discussed more fully in Chapter Three. The second section presents the computation of a similarity measure for the courses. The third section describes the cluster analysis procedure and the final section presents the discriminant analysis of the coursework patterns.

In the Coursework Cluster Analysis Model, the objects of this analysis are the courses in which students enroll. Through performing the analyses described in this chapter, empirical information can be gained concerning which coursework patterns help students to improve their learning.

Calculation of Course Mean Assessment Scores

This section describes how to relate gains in general learning to the coursework taken during the undergraduate years. To discern the contribution that a course makes to general learning, the means of the assessment scores for all the students enrolled in a particular course are chosen as the criteria measures. The mean depicts, on average, the effects of enrollment in a course. Two major steps produce the data file for use in the cluster analysis.

First, match the file containing the assessment scores to the file containing the transcript information. Ensure that each course listed for a particular student has his/her assessment scores. For example (Table 5), each student had three scores from a particular assessment. Each student score was recorded on an assessment data tape. Now looking at the data tape of student transcripts, we find that six students took Course A, three students took Course B, and five students took Course C. Course A was listed six times, Course B listed three times, and Course C listed five times. On a matrix we have three columns of assessment scores and six rows of courses from the transcripts. In each assessment score columns, there are six values for Course A corresponding to the scores of the students enrolled in that particular course. Similarly, there are three values for Course B and five values for Course C. Thus, the first step is to create a raw data file of coursework found on the students transcripts and corresponding student assessment scores.

In the second CCAM step, we compute the mean assessment scores for each course. In doing so, we also not the frequency with which the course appeared on the transcripts. To do this, you may use a procedure such as PROC MEANS in SAS or AGGREGATE in SPSSx. This outputs to a separate file. These computer procedures are listed in a sample command file in Appendix B. You probably will find it economical to include a noprint command. Datasets combining transcripts and assessment scores are large. The printouts can be voluminous and at this stage provide little valuable information. What results from this second step is an unduplicated course matrix, where each course is listed once with the corresponding mean for each assessment score, and with a count of the frequency with which the course was listed on the student transcripts (See Table 6). In this example, the resulting file is a three by four raw data matrix with Courses A, B, and C listed once. Also given are the mean for each assessment score for each course and the number of students who enrolled in the course.

Table 5. Duplicated Course File and Scores

Course	Score 1	Score 2	Score 3
Course A	4	7	4
Course A	4	7	2
Course A	5	8	5
Course A	6	9	5
Course A	5	5	1
Course A	6	6	1
Course B	4	6	2
Course B	5	7	1
Course B	3	5	3
Course C	4	8	4
Course C	6	7	3
Course C	5	9	4
Course C	7	9	5
Course C	3	7	4

Table 6. Raw Data Matrix: Unduplicated Course and Course Mean

Course	Score 1 Mean	Score 2 Mean	Score 3 Mean	Frequency
Course A	6	7	3	6
Course B	4	6	2	3
Course C	5	8	4	5

Restricting the Analysis of Low Enrollment Classes

Before moving to the next CCAM step, you must set a minimum enrollment size for the courses you wish to analyze. Thus, you need to decide on the number of students enrolling for each course desired for use in the cluster analysis. The fewer times a course appears on the transcripts, the smaller the probability that the course mean will reflect the population mean for the course. For example, if a course appeared only twice on the transcripts, the mean for any assessment score will be computed on only two student scores. By increasing the number of scores used to compute the mean, the probability that the mean will reflect the population of students who enrolled in the course increases. However, the larger the minimum frequency is, the greater the number of courses with smaller enrollments that will be excluded from the analysis. If you decide to set the lower limit to a frequency of fifteen, many courses may be dropped from the analysis due to low enrollment.

When using the CCAM for the first time, we recommend a minimum frequency of five students taking a particular course. We used this limit in the CCAM analysis of Western University (Ratcliff, 1988). Here, the assessment scores were the residuals produced from regression analyses of the nine GRE item-types on their corresponding SAT scores. The residuals were then matched to the transcript file using the previously assigned ID's, which were saved in the output residual file. This produced a file which had 3,427 courses (all courses listed on the transcripts), columns for each of the nine GRE item-types, and a column with the IDs. Calculating the means and frequencies for the courses produced a file that contained 1,088 unduplicated courses, the respective course means for each of the nine item-types, and a frequency of appearance for each course. When the criteria of 5 or more was used, the number of courses was reduced to 177.

Figure 7 indicates an actual example of the distribution of the GRE item-type residuals for the 177 courses used in the Coursework Cluster Analysis Model for Western University. Originally, the means for the GRE residuals were calculated for all courses and for courses enrolling five or more students. As the data were aggregated according to the courses enrolling five or more students, the standard deviation of these means was considerably smaller than those for all courses. This trend indicates that there are relationships between the coursework taken and the student residual scores on the assessment instrument (the GRE in this case).

Table 7. Mean Course Residuals for 177 Courses

GRE Item-types	Number of Items	Max Value	Min Value	Score Range	Residual Means	Std Error of Mean	Std Deviation
Analogy	18	3.04	-2.38	5.42	.0676	.0680	.9046
Sentence Completion	14	3.22	-1.76	4.97	.1770	.0643	.8556
Reading Comprehension	22	4.82	-3.32	8.14	.1703	.1038	1.3812
Antonyms	22	5.10	-4.23	9.33	-.0605	.1074	1.4289
Quantitative Comparison	30	7.39	-5.27	12.66	.4531	.1614	2.1470
Regular Mathematics	20	8.03	-2.50	10.53	.2035	.1310	1.7433
Data Interpretation	10	3.97	-1.94	5.91	.1861	.0794	1.0563
Analytical Reasoning	38	4.23	-5.66	9.89	.1174	.1408	1.8733
Logical Reasoning	12	2.18	-1.91	4.08	.1632	.0637	.8477
GRE Item-types:							
Minimum	10	2.18	-5.66	4.08	-.0605	.0637	.8477
Maximum	38	8.03	-1.76	12.66	.4531	.1614	2.1470
Mean	21	4.67	-3.22	7.88	.1763	.1065	1.4167
Total	186				1.4102		

Selecting the Resemblance Coefficient

Do certain courses have similar effects on general learning? To answer this question, we need a way of measuring the similarity of assessment score means among courses. Thus, the next task is to select *a similarity measure*. Some writers on cluster analysis call the similarity measure (Lorr, 1983) *the resemblance coefficient* (Romesburg, 1984). These are really two terms for the same thing, a way to measure the similarity of a field of objects according to multiple criteria. The purpose of the resemblance coefficient is to explain the similarity (or dissimilarity) of each cell to each of the other cells in the data matrix and it is expressed mathematically. There are many resemblance coefficients; each will express the similarity between courses in a slightly different way. Each coefficient is appropriate for achieving slightly different goals.

The similarity measure we recommend for the Coursework Cluster Analysis Model is the Pearson product-moment correlation coefficient, which theoretically ranges from -1.00 to 1.00. It is appropriate for use with ratio data. The correlation coefficient is produced by standardizing the assessment data across different measures. This allows for the differences in size or range of scores on different measures to be given equivalent weight. Thus, a writing sample scored on a 10 point scale will hold the same power in grouping coursework as a mathematics test scored on a 100 point scale. The resemblance coefficient indicates the similarity of courses to each other according to the residual scores for each assessment measure. In the case drawn from the Differential Coursework Patterns Project, the resemblance coefficient indicates the similarity of courses according to the residual scores on the nine GRE item-types as coded in the data matrix.

Creating the Resemblance Matrix

The resemblance matrix is formed to assess how similar, or dissimilar, one course is to another according to the assessment criteria used. The resemblance matrix is calculated by transforming the raw data matrix into ***a resemblance matrix*** using the resemblance (correlation) coefficient. The resemblance matrix contains correlations derived from the assessment criteria. In the DCP Project, the nine GRE item-types course mean residuals served as the multiple assessment measures. These correlations range in value from 1.00 to -1.00. In our example, ***the resemblance matrix*** consists of columns representing the first course in a pair, and the rows representing the second course of a pair. The resemblance coefficient (Pearson's *r*) in each cell tell us how similar the learning improvement of the students Course A was to that of Course B. Thus, the cell value represents the extent to which the attributes on the first course explain the variance in attributes on the second course. The similarity measures in the resemblance matrix are next used by the cluster analysis procedure selected to group courses according to the similarities in learning improvement of the students who enrolled in them.

In the resemblance matrix, the correlation coefficient is computed for the assessment scores' means for each pair of courses. If two courses had the same means on each of the assessment scores, the coefficient would be $r=1.00$ for that pair. The resemblance matrix has the same number of rows and columns, each representing unduplicated courses found on the students transcripts. For example, if a data set contained six courses, a six by six data matrix is computed by calculating the correlation coefficient of each course with each of the other five courses. As Table 8 shows, Course A is very similar to Course E; the assessment scores' means correlated at $r = .90$. Course C, on the other hand, is dissimilar to Course D ($r = -.25$).

Table 8. Correlation Coefficient as Similarity Index for Six Unduplicated Courses

Course	Course A	Course B	Course C	Course D	Course E	Course F
Course A	1.00	.57	.35	-.01	.90	.74
Course B	.57	1.00	.49	.31	.51	.70
Course C	.35	.49	1.00	-.25	.30	.53
Course D	-.01	.31	-.25	1.00	.01	-.08
Course E	.90	.51	.30	.01	1.00	.85
Course F	.74	.70	.53	-.08	.85	1.00

Choosing a Method of Cluster Analysis

Using the resemblance matrix, courses enrolling students with comparable levels and types of improvement can be associated with one another. But, noting the similarities between any two courses tells little about the similarities between many courses. Undergraduate education is a progressive, cumulative experience where curricular and extracurricular experiences build upon one another to effect student learning. Cluster analysis allows us to group those patterns of coursework together which enrolled students who showed large improvement in one or more of the assessment criteria. By clustering courses according to their effects, we can ask some important questions: Do several courses contribute together to influence learning? Do sequences of courses consistently contribute to areas of general learning? As you can see, coursework patterns are the specific focus of the CCAM analysis.

Do different patterns of coursework have unique contributions? While it is theoretically possible to examine the resemblance matrix course by course to determine which courses contribute in a similar manner to the assessment scores, the immensity of this task examining perhaps over two hundred courses would keep several researchers busy for some time. Cluster analysis lets us accomplish this task quickly and with relative ease.

Cluster analysis forms groups, or clusters, of courses according to a measure of similarity (or dissimilarity). The courses are first treated as each being a separate cluster (all courses are dissimilar). For example, if the number of unduplicated courses on a group of student transcripts was 200, then the beginning number of clusters is 200. In each successive clustering, courses are grouped according to the similarity measure. In the first step, all course coefficients are compared and the two courses (clusters) most similar are joined as a new cluster. The similarity of the new cluster to the other courses is then computed. The computation (or clustering) method selected using with the CCAM analysis is ***the unweighted pair-group method using arithmetic means*** (UPGMA). This clustering method is readily available in the statistical procedure was CLUSTER in SPSSx. The clustering computation is performed by taking the similarity measures of the two courses joined as a cluster and averaging their similarities with each of the remaining clusters. For example, in Table 9, the two courses with the highest correlation coefficient, Course A and Course E (r = .90), are the most similar and join to form the first cluster. Their similarity coefficients with each of the other four courses are averaged and become the similarity for the new cluster, Cluster (A,E), with each of the other four clusters (still one course each). As Course A was correlated with Course B at r = .57 and Course E was correlated with Course B at r = .51, the new similarity measure for Cluster (A,E) and Cluster B would be ((.57 + .51)/2) or .54 (see Table 9). This same procedure would be applied to determine the similarity measures for Cluster (A,E) and each of the remaining three clusters. This reduces the similarity matrix by one row and one column leaving five remaining clusters. A second step would repeat the process, reducing the matrix again by one row and one column, leaving the matrix with four remaining clusters. The number of steps needed would be one less than the number of courses originally in the similarity matrix, with the final clusters joining to form one cluster that includes all of the courses.

Table 9. Recalculated Similarity Coefficients after Course A and Course E are joined to Form Cluster (A,E)

Cluster	Cluster (A,E)	Cluster B	Cluster C	Cluster D	Cluster F
Cluster (A,E)	1.00	.54	.32	.00	.80
Cluster B	.54	1.00	.49	.31	.70
Cluster C	.33	.49	1.00	-.25	.53
Cluster D	.00	.31	-.25	1.00	-.08
Cluster F	.80	.70	.53	-.08	1.00

Deciding on the Number of Clusters

The decision on the number of clusters an institutional researcher chooses is arbitrary. The clusters range from each course being a separate cluster (no two courses are similar) to only one cluster in which all courses are treated as similar. 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. Obviously, if all courses are treated as similar (one cluster), little information is gained from a cluster analysis. On the other hand, if no two courses are treated as similar, again little information is gained on coursework patterns. SPSSx produces a dendrogram that plots the joining of clusters on a relative distance scale that is helpful in determining the number of clusters desired.

Table 10 provides an example of an SPSSx dendrogram produced for a data file of 25 courses. A six cluster solution has been identified to facilitate interpretation. The recalculated distances (similarities) are relative to the final cluster formed, as relative distance increases, clusters that are more dissimilar are joined. In the sample dendrogram, the cluster labeled #1 was formed at a relative distance of six, as was the cluster labeled #2. Cluster #3 did not form until a distance of eight while cluster #5 demonstrates greater similarity, forming at a distance of four. Cluster #4 is relatively dissimilar to the other clusters, not joining with any clusters until distance sixteen and cluster #1 is the most dissimilar to the other clusters, not combining until the final stage. If we were to move to a 5-cluster solution, clusters #5 and #6 would join next at a relative distance of 10 and a four cluster solution would join clusters #2 and #3 at relative distance eleven.

Note the column labeled SEQ. This is the sequence number of the course as it is read into the active file. The sequence number will be used through out the remaining analysis and is alternately called the case number. For example, in the dendrogram below, BIO 201 has a case sequence of 5.

Table 10. Sample Course Cluster Dendrogram

```
================================================================================
DENDROGRAM USING AVERAGE LINKAGE (BETWEEN GROUPS)

                             RESCALED DISTANCE CLUSTER COMBINE

  C A S E         0         5        10        15        20        25
LABEL        SEQ  +---------+---------+---------+---------+---------+

Bio     201    5   -+
Bio     202    6   -+
Chem    113    8   -+-+
PhysTh  103   18   -+ |
Chem    111    7   -+ +-------+
Psy     102   21   -+ |       |
Bio     101    3   -+-+       +------#1-----------------------------+
Bio     102    4   -+ |       |                                     |
Psy     101   20   ---+       |                                     |
Math    111   15   -----------+                                     |
Anthro  104    1   ---+-------+                                     |
Soc     101   22   ---+       +------#2-+                           |
Econ    121   10   -----------+         +-------------+             |
Engl    219   12   ---+-----------+     |             |             |
Thea    160   24   ---+           +--#3-+             |             |
Pol     101   19   ---------+-----+                   +-------------+
W&R     106   25   ---------+                         |
MA      111   17   -----------------+#4-----------+   |
Thea    100   23   -----------------+             |   |
CSci    110    9   ---+---+                       +---+
GBus    303   13   ---+   +---------#5+           |
AC      105    2   ---+---+           |           |
Mktng   212   16   ---+   |           +-----------+
Math    105   14   -------+           |
Econ    122   11   -----------------#6+
================================================================================
```

The Agglomeration Schedule

Another helpful printout from SPSSx is the Agglomeration Schedule. It indicates the similarity coefficient at which the clusters are formed. Table 11 is an agglomeration schedule derived from the 25 course data in the sample dendrogram (See Table 10). The clusters in the dendrogram were named Cluster #1 through Cluster #6 for the sake of convenience. The clusters in the Table 10 present the case number (or as mentioned above, the clustering sequence number) for each course. SPSSx names a new cluster beginning with the lowest sequence number. Recall that each course begins as a separated cluster. Case #5 is also cluster #5 at the start. In subsequence stages of clustering, cases are joined according to their similarity. For example, case #5 (BIO 201) joins with case #6 (BIO 202) in stage one and the two courses become cluster 5 in the schedule until it is merged with a lower course sequence number. In this instance, the course that was third in the course sequence merges with 5 at stage 10. The schedule prints the stages of clustering and the similarity coefficient at which the clusters were joined. Recall that the clustering is done in stages and each stage represents a cluster formation. In order to locate a six cluster solution, count up from the last stage six stages. Table 9 reported a correlation similarity coefficient for the final merged cluster of a six-cluster solution; the coefficient, $r = .54$, demonstrates a reasonable similarity for the two clusters joined according to the assessment scores. All clusters formed previously will have a higher correlation coefficient. Again, the degree of dissimilarity tolerated is up to the researcher. As will be seen in subsequent steps of the CCAM, lower similarity coefficients can be useful in relating a group of courses to one particular assessment score.

Table 11. Agglomeration Schedule Using Average Linkage (between groups)

Stage	Clusters Combined Cluster 1	Clusters Combined Cluster 2	Coefficient	Stage Cluster Appears Cluster 1	Stage Cluster Appears Cluster 2	Next Stage
1	5	6	1.00	0	0	6
2	8	18	.99	0	0	3
3	7	8	.99	0	2	5
4	3	4	.99	0	0	7
5	7	21	.99	3	0	6
6	5	7	.98	1	5	10
7	3	20	.94	4	0	10
8	12	24	.92	0	0	18
9	2	16	.92	0	0	13
10	3	5	.92	7	6	16
11	1	22	.91	0	0	17
12	9	13	.91	0	0	14
13	2	14	.84	9	0	14
14	2	9	.80	13	12	20
15	19	25	.74	0	0	18
16	3	15	.72	10	0	24
17	1	10	.69	11	0	21
18	12	19	.60	8	15	21
19	17	23	.54	0	0	22
20	2	11	.49	14	0	22
21	1	12	.45	17	18	23
22	2	17	.18	20	19	23
23	1	2	.09	21	22	24
24	1	3	-.29	23	16	0

Storing and Saving the Work

At this point in the CCAM analysis, it is a good idea to save your work. In fact it is necessary save part of the printout electronically for use in the subsequent analysis. SPSSx does not currently save cluster membership in an active file. The portion that needs to be saved is titled Cluster Membership of Cases. Table 12 is a sample printout from the 25 course data set. The listing must be edited for input as a raw data file, removing all titles, page numbers, and headings. With large data sets, the printout will be lengthy with numerous page breaks. All page breaks must be taken out producing a final raw data file consisting of course names and cluster membership. Be sure to note the locations of each cluster in the raw file to instruct SPSSx about the locations of the clusters. The number of different cluster solutions printed out can be determined by the researcher. We found in the DCP Project research that the normal range is from a five cluster solution to a twenty-five solution. The printout in Table 12 shows the range from a five-cluster solution to a 15-cluster solution. Most text editors (such as XEDIT in CMS) are capable of this editing the Cluster Membership printout.

Table 12. Hierarchical Cluster Analysis

Label		Case	Number of Clusters 15	14	13	12	11	10	9	8	7	6	5
Anthro	104	1	1	1	1	1	1	1	1	1	1	1	1
AC	105	2	2	2	2	2	2	2	2	2	2	2	2
Bio	101	3	3	3	3	3	3	3	3	3	3	3	3
Bio	102	4	3	3	3	3	3	3	3	3	3	3	3
Bio	201	5	3	3	3	3	3	3	3	3	3	3	3
Bio	202	6	3	3	3	3	3	3	3	3	3	3	3
Chem	111	7	3	3	3	3	3	3	3	3	3	3	3
Chem	113	8	3	3	3	3	3	3	3	3	3	3	3
CSci	110	9	4	4	4	4	2	2	2	2	2	2	2
Econ	121	10	5	5	5	5	4	4	4	1	1	1	1
Econ	122	11	6	6	6	6	5	5	5	4	4	4	2
Engl	219	12	7	7	7	7	6	6	6	5	5	5	4
GBus	303	13	8	8	4	4	2	2	2	2	2	2	2
Math	105	14	9	9	8	2	2	2	2	2	2	2	2
Math	111	15	10	10	9	8	7	7	3	3	3	3	3
Mktng	212	16	2	2	2	2	2	2	2	2	2	2	2
MA	111	17	11	11	10	9	8	8	7	6	6	6	5
PhysT	103	18	3	3	3	3	3	3	3	3	3	3	3
Pol	101	19	12	12	11	10	9	9	8	7	5	5	4
Psy	101	20	3	3	3	3	3	3	3	3	3	3	3
Psy	102	21	3	3	3	3	3	3	3	3	3	3	3
Soc	101	22	13	1	1	1	1	1	1	1	1	1	1
Thea	100	23	14	13	12	11	10	10	9	8	7	6	5
Thea	160	24	7	7	7	7	6	6	6	5	5	5	4
W&R	106	25	15	14	13	12	11	9	8	7	5	5	4

The label identifies the course. The case number is the order in which the data set was read. The numbers immediately following the heading CASE represent the number of clusters. The heading 15 denotes a 15 cluster solution. The numbers under the heading of 15 denote the cluster in which each course is grouped in a 15 cluster solution. The next step in the CCAM analysis calls for joining the data set resulting from the editing of the cluster membership with the original course means data set. Joining this two data sets is a prerequisite to perform the discriminant analysis which follows (see Appendix B). Table 13 is the final raw data file for input into an SPSSx file.

Table 13. Sample Final Raw Data File

Anthro	104	1	1	1	1	1	1	1	1	1	1	1	1
AC	105	2	2	2	2	2	2	2	2	2	2	2	2
Bio	101	3	3	3	3	3	3	3	3	3	3	3	3
Bio	102	4	3	3	3	3	3	3	3	3	3	3	3
Bio	201	5	3	3	3	3	3	3	3	3	3	3	3
Bio	202	6	3	3	3	3	3	3	3	3	3	3	3
Chem	111	7	3	3	3	3	3	3	3	3	3	3	3
Chem	113	8	3	3	3	3	3	3	3	3	3	3	3
CSci	110	9	4	4	4	4	2	2	2	2	2	2	2
Econ	121	10	5	5	5	5	4	4	4	1	1	1	1
Econ	122	11	6	6	6	6	5	5	5	4	4	4	2
Engl	219	12	7	7	7	7	6	6	6	5	5	5	4
GBus	303	13	8	8	4	4	2	2	2	2	2	2	2
Math	105	14	9	9	8	2	2	2	2	2	2	2	2
Math	111	15	10	10	9	8	7	7	3	3	3	3	3
Mktng	212	16	2	2	2	2	2	2	2	2	2	2	2
MA	111	17	11	11	10	9	8	8	7	6	6	6	5
PhysTh	103	18	3	3	3	3	3	3	3	3	3	3	3
Pol	101	19	12	12	11	10	9	9	8	7	5	5	4
Psy	101	20	3	3	3	3	3	3	3	3	3	3	3
Psy	102	21	3	3	3	3	3	3	3	3	3	3	3
Soc	101	22	13	1	1	1	1	1	1	1	1	1	1
Thea	100	23	14	13	12	11	10	10	9	8	7	6	5
Thea	160	24	7	7	7	7	6	6	6	5	5	5	4
W&R	106	25	15	14	13	12	11	9	8	7	5	5	4

Discriminant Analysis of Coursework Patterns

The purpose of a discriminant analysis is two-fold. First, the discriminant analysis provides a secondary validation of the cluster analysis. It provides a classification analysis determining the percent of courses correctly grouped and probabilities for courses being in other groups. Second, the discriminant analysis examines the relationship of the cluster groupings to the assessment scores.

A discriminant function is a linear relationship of discriminating variables. In this case the relationship is that of the mean assessment scores to the groupings of cases. In the

DCP Project research, the cases were course groupings. As its name implies, the purpose is to discriminate between the groups, i.e. assess the mutual exclusivity of the groups. The functions are derived to provide the maximum distance between means of the groupings while remaining uncorrelated to each other. The maximum number of functions that can be derived is the number of groupings minus one (g-1) or the number of discriminating variables, which ever is less. A detailed explanation of the discriminant procedure will not be presented (for those interested in a detailed presentation please refer to Haggerty, 1975; Norusis, 1985). The discriminant analysis was performed by the procedure DISCRIMINANT in SPSSx (See Appendix B for job control codes).

To classify a course as belonging to a group, the classification coefficients of the groups is used to check if the classification score for a particular course is the highest possible. Suppose that a course is in group #1, the classification coefficients for each of the groups would be applied to the mean assessment scores for the course. If the coefficients that produced the highest classification score for the course were the coefficients from group #1, the course would have been correctly classified. If, in fact, the coefficients from group #2 produced the highest score, the group was then misclassified. The prior probabilities that a course belongs in a group used by the DCP is the proportion of the courses in the group. Table 14 is the printout of the classification data from the 25 course data set shown in the cluster analysis using a 6 cluster solution. Those cases marked with asterisks were misclassified. There are no asterisks in this printout as all courses were correctly classified.

Table 14. Sample SPSSx Printout of Classification Analysis

Case Number	Mis. Val	Sel	Actual Group	Highest Group	Probability P(D/G)		P(G/D)	2nd Highest Group	P(G/D)
1			1	1	0.8617	1	1.0000	4	0.0000
2			2	2	0.5371	0	.9998	3	0.0001
3			3	3	0.2276	1	1.0000	4	0.0000
4			3	3	0.5221	1	1.0000	4	0.0000
5			3	3	0.8490	1	1.0000	4	0.0000
6			3	3	0.8490	1	1.0000	4	0.0000
7			3	3	0.7030	1	1.0000	4	0.0000
8			3	3	0.9732	1	1.0000	4	0.0000
9			2	2	0.6003	0	.9999	4	0.0001
10			1	1	0.4630	0	.9457	4	0.0540
11			4	4	1.0000	0	.9995	1	0.0003
12			5	5	0.2516	1	1.0000	1	0.0000
13			2	2	0.4538	1	1.0000	4	0.0000
14			2	2	0.4228	1	1.0000	4	0.0000
15			3	3	0.3269	0	.9963	2	0.0024
16			2	2	0.3593	0	.9505	4	0.0495
17			6	6	0.6296	1	1.0000	3	0.0000
18			3	3	0.9349	1	1.0000	4	0.0000
19			5	5	0.3337	1	1.0000	2	0.0000
20			3	3	0.6313	0	1.0000	2	0.0002
21			3	3	0.6614	1	1.0000	4	0.0000
22			1	1	0.4545	1	1.0000	4	0.0000
23			6	6	0.6296	1	1.0000	3	0.0000
24			5	5	0.3957	1	1.0000	2	0.0000
25			5	5	0.5859	1	1.0000	1	0.0000

The first column identifies the case sequence number, which will be the same as the sequence number in the cluster analysis. The second column identifies the number of missing values. The third column, marked SEL, is used for a select variable. The DCP does not currently use a select variable. The fourth column identifies the original group membership. If a course is misclassified, asterisks will appear in this column (see SPSSx manual for the remaining identifications).

As is shown in Table 15 there were no misclassifications in the six cluster solution. If misclassifications did occur, the percentage of misclassifications for any given cluster generally should not exceed fifty percent. If more than 50 percent of the courses are misclassified, the cluster cannot reliably be related to the functions.

Table 15. Sample SPSSx Analysis of Coursework Classification Results

Actual Group	Number of Cases	1	2	3	4	5	6
GROUP 1	3	3 100.00%	0 .00%	0 .00%	0 .00%	0 .00%	0 .00%
GROUP 2	5	0 .00%	5 100.00%	0 .00%	0 .00%	0 .00%	0 .00%
GROUP 3	10	0 .00%	0 .00%	10 100.00%	0 .00%	0 .00%	0 .00%
GROUP 4	1	0 .00%	0 .00%	0 .00%	1 100.00%	0 .00%	0 .00%
GROUP 5	4	0 .00%	0 .00%	0 .00%	0 .00%	4 100.00%	0 .00%
GROUP 6	2	0 .00%	0 .00%	0 .00%	0 .00%	0 .00%	2 100.00%

Percent of Groups Correctly Classified 100 %

Correlating Course Mean Assessment Scores and Discriminant Functions

The pooled within-groups correlation coefficients are used to examine the relationship between the discriminating variables (mean assessment scores) and the discriminant functions. To compute the correlation, discriminant function scores are computed over all courses for each function. A function's scores are then correlated with each of the discriminating variables, revealing the strength and direction of the relationship with each of the assessment scores. Table 16 shows the SPSSx printout for the pooled within-group correlations for the 6-cluster solution of the 25 course data set and the mean assessment scores.

Table 16. Pooled Within-groups Correlations Between Discriminating Variables and Canonical Discriminant Functions

	Function 1	Function 2	Function 3	Function 4	Function 5
MRAR	.5799 *	-.1918	-.0871	.3335	.2803
MRLR	-.0989	.6320 *	.2430	.0078	.4251
MRANT	-.2817	.5398 *	.0990	.1684	.2812
MRDI	-.0687	-.2891 *	-.1139	.2440	-.0906
MRRM	.4154	-.4163	.5212 *	-.2944	-.1065
MRQC	.4708	.0333	.0777	-.5787 *	-.1739
MRANA	-.1590	.0470	-.2763	.3772 *	-.0583
MRSC	-.2110	.4449	-.1046	-.0666	.6407 *
MRRD	.1636	-.0746	-.1361	-.1036	.5332 *

The correlations of the mean assessment scores establishes the strength of the relationship, either positive or negative, between a function and each of the assessment scores. The asterisks denote the highest correlation for an assessment score's means and the functions. The CCAM has used the cutoff point for a significant correlation of .50, positive or negative. Again it will be the decision of the researcher to decide a satisfactory level of a correlation.

Explanatory Power of the Discriminant Functions

The pooled within-group correlations established relationships between the discriminant functions and the assessment scores. The researcher must now decide how many functions provide meaningful information. Each discriminant function explains a certain proportion of the variation in mean assessment scores.

Discriminant functions with strong explanatory power, "good discriminant functions," have large between-cluster variability and low within-cluster variability (Haggerty, 1975; Norusis, 1985). The eigenvalues of Table 17 present the ratio of between-group to within-group sums of squares of the residuals. Large eigenvalues are associated with the discriminant functions that most contribute to explaining variability in mean assessment scores.

Wilk's Lambda is the ratio of the within-group sum of squares to the total sum of the squares. It represents the proportion of the total variance in the discriminant function values not explained by differences among cluster groups. Wilk's Lambda serves as a test of the null hypothesis that there is no difference in the mean residuals of a coursework cluster means and the mean residual scores of the coursework in the total sample.

Thus, the eigenvalues and canonical correlations indicate the extent to which each discriminant function contributes to our understanding of the variability in coursework mean assessment scores. Lambda test the null of the differential coursework hypothesis for each discriminant function. The percent of variance accounted for by a function and the probability of error is also shown in Table 17. The CCAM uses the percent of variance accounted for by a function as the criteria for inclusion in the analysis due to its wide spread interpretability. The first function created will account for a greater amount of the variability than the second function created. The second function will account for more variability than the third. The minimum level set for the CCAM analysis is 5% of the variance. While some functions may account for less than five percent of the variance and still be considered significant, the additional information provided is weak compared to those functions that account for the majority of the variability. As can be seen in Table 17, the first four functions account for over 98 percent of the variance.

Table 17. Sample SPSSx Printout of Eigenvalues

Function	Eigen-Value	Percent of Variance	Cumulative Percent	Canonical Correlation	Wilk's Lambda	Degrees Freedom	Signifi-cance
0					0.0004	45	0.0000
1*	15.6860	54.62%	54.62%	0.9696	0.0062	32	0.0000
2*	6.9085	24.06%	78.68%	0.9346	0.0490	21	0.0004
3*	3.2847	11.44%	90.11%	0.8756	0.2100	12	0.0116
4*	2.4656	8.59%	98.70%	0.8435	0.7279	5	0.3872
5*	0.3739	1.30%	100.00%	0.5217			
0							

The Relationship of Coursework Clusters, Functions, and Assessment Scores

The pooled within-group correlations revealed the relationships between the functions and the assessment score means. It does not reveal a relationship of the clusters to the assessment score means, which is the next step in the discriminant analysis. Discerning the relationship between the individual clusters and the assessment scores is done by evaluating the group centroids and then using the pooled within-group correlations. The centroids are a result of the function scores for the courses in the clusters and are essentially the means of the discriminant scores for a cluster on a particular function. If four functions were computed, each cluster would have four centroids. The relationship between the assessment scores and the clusters are established by noting the correlation of an assessment score to a function and the size of the centroid. If a cluster had a high positive centroid on a function and the function was positively correlated to one of the assessment scores, the cluster contributes positively to that assessment score. If a centroid proved to be large and negative on a function, and that function was positively correlated to an assessment score, the cluster does not contribute to the assessment score. On the other hand, if a cluster's centroid was large and negative on a function which was strongly and negatively correlated to an assessment score, the cluster contributes to the learning measured by the assessment score. An SPSSx printout of the evaluation of functions at the group centroids is presented in Table 6k. The DCP used a centroid value greater than 1.00, negative or positive, in the evaluation of GRE residuals. An example of the interpretation of the relationship between assessment scores, functions, and clusters is presented next.

As shown in Table 16, MRAR was correlated positively to Function 1 (r=.58). Looking under Function 1 in Table 18, group # 1 had a high negative centroid, demonstrating that those courses in cluster #1 did little to enhance learning in the area assessed by MRAR. This was not the case for group #6. The courses in group six produce a high positive centroid for Function 1 indicating that these courses enhanced the learning in the area measured by MRLR. On the other hand, MRQC was negatively correlated to Function 4 (r=-.57). Checking the groups centroids for Function 4 in Table 18 reveals that group #3 had a high negative value, which would indicate that courses in this group contributed to the area of learning measured by MRQC. The negative relationship of MRQC to Function 4 indicates that moving down the continuum of Function 4 improves the scores of MRQC.

Table 18. SPSSx Printout of Canonical Discriminant Functions Evaluated at Group Means

Group	Function 1	Function 2	Function 3	Function 4	Function 5
1	-2.2960	-.6562	-3.7081	1.4269	-.0999
2	.8883	-3.0953	-.0858	-.6959	-.2968
3	1.1635	2.7311	.8029	-1.4518	-.6798
4	.1768	-.4383	-1.2162	1.9125	-2.4909
5	-4.9763	-2.7835	2.4776	2.4319	.4502
6	5.2700	.8532	-2.5852	1.0385	4.6359

Summary

This chapter gave detailed procedures from using the CCAM to link assessed improvement in student learning with specific patterns of coursework. By linking what student learning experiences with student learning, specific actions can be taken to improve and reform the curriculum. Colleges and universities can test the efficacy of their general education curriculum (Ratcliff, 1990c), determine if transfer students from the community college benefit from the same coursework as those who began their undergraduate education at the baccalaureate degree granting institution (Ratcliff and Jones, 1991), or determine if the same core coursework effects the learning of different groups of students in different ways. The results of the analysis can be used to examine instructional strategies, grading practices and common expectations of instructors teaching courses within a cluster pattern. In short, while the CCAM analytic procedures are detailed and complete, they do provide the promise for substantive and specific improvement of undergraduate education.

APPENDIX A
CATALOG STUDIES

The *purpose* of a catalog study is two-fold: (1) to determine an institution's degree requirements and general education requirements, and (2) to trace the pattern of course changes (additions, deletions, development of equivalencies and cross listings) for the courses taken by the students in the institutional sample.

In doing a catalog study, the first step is to create an outline containing the following three frames:

1. Course by catalog year matrix (spreadsheet)
2. Narrative of course equivalencies, cross-listings, additions, and deletions
3. Course prerequisite matrix (spreadsheet)

Course by Catalog Year Matrix

1. Download onto a matrix the courses from the student transcript data, using department codes, numbers, and titles. Exclude courses with only one student in them.

   NOTE: If titles are missing, the current catalog is the starting point to get course titles. The aim is to get 95% of all the courses being examined. Thus, the number of catalogs studies will vary, depending on how many are necessary to use to get coverage of 95% of the courses taken.

2. For each catalog year being studies, put a Y (for Yes) or an N (for No) to indicate if the course was listed in a particular catalog.

3. If there is no title with the course number, write down the title that is found the first time the course number appears in the catalogs being studies. For example, using catalogs from 1979 through 1986, look for AC 201. If it first appears in the 1981-82 catalog as "Principles of Accounting," write this down as the course title.

DEPT	CNUM	TITLE	EQIV	80-81	81-82	82-83	83-84
AC	201	Principles of Accounting	N	Y	Y	Y	

4. If the course has the same number but changes its title, write down the title change and indicate the year the title changed.

DEPT	CNUM	TITLE	EQIV	80-81	81-82	82-83	83-84
AC	201	Principles of Accounting	N	Y			
AC	201	Accounting Principles			Y	Y	

Narratives of Course Equivalencies, Crosslistings, Additions, and Deletions

After doing the course by catalog year matrix, use the information on it to fill out the narrative forms. (See attached sample forms.) The forms should be filled out thusly.

1. <u>Course Equivalencies (including Crosslistings)</u>:

   Write down the original course, the course to which it is now equivalent or with which it is crosslisted, and the catalog year the equivalency or crosslisting first appeared.

MATH 101 = MATH 201	83-84
DRAM 101 = ART 101	82-84

2. <u>Course Additions and Deletions</u>

   Write down the course and the year of the catalog in which it first appeared or from which it was deleted.

Course Prerequisite Matrix

1. Using a spreadsheet, list each course (with prerequisites) alphabetically and numerically by department name and course number.

2. List the year(s) each prerequisite was required and list each prerequisite.

3. Calculate the number of prerequisites for course being studied.

DEPT	CNUM	YR	PREREQ#1	PREREQ#2	PREREQ#3	CALC
MATH	101	81-82	MATH 090	MATH 089		2
MATH	101	82-83	MATH 090			1

4. Ignore any comments about consent of instructor being a prerequisite.

5. In case of prerequisites linked by "or," list all prerequisites in PREREQ#1 column.

DEPT	CNUM	YR	PREREQ#1	PREREQ#2	PREREQ#3	CALC
MATH	101	81-82	MATH 090 MATH 095 MATH 099			1

Definitions: Equivalencies, Additions, Deletions

Equivalencies

1. If the course number/title changes but the course title/number does not, check the two course descriptions. If they are the same or very similar, count the course with the changed title/number as an equivalency.

2. If the department code changes but the number and title do not, treat the course as an equivalent. Also make a note that the department changed its name in a particular year.

3. If the course number an/or title change(s), check the two course descriptions. If the descriptions are the same or very similar, treat the courses as equivalents.

Additions/Deletions

1. If the course number an/or title change(s), check the two course descriptions. If they are different, treat the original course as a deletion and the current course as an addition.

2. If a course appears for the first time in a catalog, the course is an addition.

3. If a course which appeared in the previous catalog does not appear in the next one, count the course as a deletion.

Course Equivalencies

Same Department						Different Department (Crosslistings)			
DEPT	CNUM =	DEPT	CNUM	YR	EQUIV	DEPT	CNUM/DEPT	CNUM	YR EQUIV OCCURRED
AC	201	AC	301	83-84		ART	202/DRAM	202	85-86

Course Additions/Deletions

Additions		Deletions	
MATH 101	83-84	MATH 101	83-84

APPENDIX B

SPSS COMMANDS FOR THE CCAM PROCEDURES

This appendix presents the job control language required to perform the Coursework Cluster Analytic Model procedures. The SPSS commands to run the analysis of coursework patterns are:

```
//V03XXXXX JOB
// EXEC SPSSX,PARM=3000K
//IN1 DD DSN=MEN.U69640.HTS.DEMO.RAW,DISP=(OLD,KEEP),
// VOL=REF=MEN.U69640.HTS.LIB
//IN2 DD DSN=MEN.U69640.HTS.TRANS.RAW,DISP=(OLD,KEEP),
// VOL=REF=MEN.U69640.HTS.LIB
//IN3 DD DSN=MEN.U69640.HTS.GRE.RAW,DISP=(OLD,KEEP),
// VOL=REF=MEN.U69640.HTS.LIB
//DEMO DD DSN=MEN.U69640.HTS.DEMO.SPSS,DISP=(NEW,KEEP),
// VOL=REF=MEN.U69640.HTS.LIB,SPACE=(TRK,(20,20),RLSE)
//CRS DD DSN=MEN.U69640.HTS.EQCRS.SPSS,DISP=(NEW,KEEP),
// VOL=REF=MEN.U69640.HTS.LIB,SPACE=(TRK,(40,40),RLSE)
//GRE      DD      DSN=MEN.U69640.HTS.GRE.SCORE.SPSS,DISP=
      (NEW,KEEP),
// VOL=REF=MEN.U69640.HTS.LIB,SPACE=(TRK,(40,40),RLSE)
//RES      DD      DSN=MEN.U69640.HTS.GRE.RESID.SPSS,DISP=
      (NEW,KEEP),
// VOL=REF=MEN.U69640.HTS.LIB,SPACE=(TRK,(20,20),RLSE)
//CRS1     DD      DSN=MEN.U69640.HTS.CRS.RESID.SPSS,DISP=
      (NEW,KEEP),
// VOL=REF=MEN.U69640.HTS.LIB,SPACE=(TRK,(70,40),RLSE)
//MRES     DD      DSN=MEN.U69640.HTS.CRS.MRESID.SPSS,DISP=
      (NEW,KEEP),
// VOL=REF=MEN.U69640.HTS.LIB,SPACE=(TRK,(40,40),RLSE)
```

Please Note: The job control language above may differ from the language your facility uses. Some changes may have to be made. The language is left in so that the reader can see what files are created and needed. In the job control language above, those files identified as raw are the input (the text files on tapes) for transcript, demo, and GRE system files.

The following commands create a system file from a raw data tape which contains two records for each sample member. The tape contains the back-ground characteristics for the sample as well as the SAT scores. The data are arranged in columns. A codebook should be provided to that reports the variables and the columns the variables are located in. In the example below, the variables are listed first and then the column(s) the data are located in. When the letter A appears in parentheses, it is an indication that the data should be read as alpha-numeric (a string variable). The /1 indicates that the information forthcoming is located on the first record.

```
DATA LIST FILE=IN1 RECORDS=2
          /1 ID 1-4
             BRTHDATE 5-10
             SEX 11 (A)
             RACE 12
             EOPHEOP 13
             MATDATE 14-17
             SATV 18-20
             SATM 21-23
             PFEARN 24-29
             CRATMPT 30-35
             CREARN 36-41
             QPEARN 42-47
             GPA 48-52
             XFRCR 53-58
             TOTDEGCR 59-64
          /2 MAJOR1 1-4
             MAJOR2 5-8
             MAJ1DPT 9-10
             MAJ2DPT 11-12
             MAJ1SCH 13-14
             MAJ2SCH 15-16
             MINOR1 17-18
             MINOR2 19-20
             MINOR3 21-22
```

```
            CONC1 23-24
            CONC2 25-26
            CONC3 27-28
COMPUTE SATT=SATM + SATV
FORMATS ID (F4.0)
```

The following adds labels to the variables.

```
VARIABLE LABELS  ID 'PARTICIPANT CODE #'
            BRTHDATE 'BIRTHDATE (YR-MO-DAY)'
            SEX 'SEX'
            RACE 'RACE'
            EOPHEOP 'RECEIPT OF EOP AND/OR HEOP AID'
            MATDATE 'DATE OF MATRICULATION (YR-MO)'
            SATV 'SAT-VERBAL'
```

The following adds values of the variables to the system file.

```
VALUE LABELS  RACE (0) NOT INDICATED
             (1) 'BLACK/NON-HISPANIC'
             (2) AMERICAN INDIAN
             (3) ASIAN
         MINOR1 MINOR2 MINOR3
          (01)RECREATION MINOR
          (02)OUTDOOR RECR MINOR
          (03)ART MINOR
          (04)SPEECH MINOR
          (05)THEATRE MINOR
          (06)DANCE MINOR
          (07)MATHEMATICS MINOR
```

The following saves the file and maps (lists) the variables.

```
SAVE OUTFILE=DEMO/MAP
```

Reading the second raw tape. Each subject has only one record. The data tape contains the necessary transcript information.

```
DATA LIST FILE=IN2 /
        SESSION 10-12 (A)
        DEPTNUM 13-14 (A)
        CRSNUM 15-17 (A)
        SECTION 18-19
        REPEAT 20 (A)
        GRDTYPE 21 (A)
        CREDHRS 22-26 (2)
        GRADE 27-28 (A)
        INSTRID 29-37
        ID 38-41
FORMATS  ID (F4.0) INSTRID (F9.0)
```

The following commands create string (Alpha-numeric) variables. This sample tape contained the departments listed by departmental code. For clarity, departmental names are used in the analysis. Other string variables are created to use in the conversion of equivalent courses in order to keep original course identifiers. It is essential that consistency in length of string variables is maintained.

```
STRING DEPT (A7)
STRING COURSE (A10)
STRING CRSN (A3)
STRING EDEPTNUM (A2)
STRING EQCRS (A10)
STRING EDEPT (A7)
```

The following converts the departmental code to departmental name.

```
IF (DEPTNUM=' 1') DEPT='EIL    '
IF (DEPTNUM=' 2') DEPT='BioChe '
IF (DEPTNUM=' 3') DEPT='Bio    '
IF (DEPTNUM=' 4') DEPT='Chem   '
IF (DEPTNUM=' 5') DEPT='Thea   '
```

The following concatenates the departmental name and course number to produce the course (i.e. Chem 213)

```
COMPUTE COURSE = CONCAT(DEPT,CRSNUM)
```

Add labels and values if desired.

```
VARIABLE LABELS  SESSION 'YEAR AND SEMESTER COURSE WAS
    TAKEN'
    DEPTNUM 'DEPARTMENT # IN WHICH COURSE WAS TAKEN'
    DEPT 'ORIGINAL DEPT'
    COURSE 'ORIGINAL COURSE NAME'
    SECTION 'SECTION # OF COURSE TAKEN'
    REPEAT 'REPEATED COURSE INDICATOR'
    GRDTYPE 'GRADE OPTION'
    CREDHRS '# CREDIT HRS FOR WHICH COURSE IS OFFERED'
    GRADE 'COURSE GRADE'
    INSTRID 'ID# OF COURSE INSTRUCTOR'
    ID 'STUDY PARTICIPANT CODE #'

VALUE LABELS  SESSION '826' 'FALL   1982'
              '832' 'SPRING 1983'
              '834' 'SUMMER 1983'
              '836' 'FALL   1983'
```

These next commands use the string variables created previously to start the formation of equivalent courses.

```
COMPUTE EDEPTNUM=DEPTNUM
COMPUTE CRSN=CRSNUM
```

The following is an example of equivalent course formation. The department code is used for convenience (less typing).

```
IF (DEPTNUM='80' AND CRSNUM='125') CRSN='105'
IF (DEPTNUM='80' AND CRSNUM='126') CRSN='106'
IF (DEPTNUM=' 6' AND CRSNUM='304') EDEPTNUM='40'
IF (DEPTNUM='25') EDEPTNUM='23'
IF (DEPTNUM=' 7' AND CRSNUM='209') CRSN='109'
IF (DEPTNUM=' 7' AND CRSNUM='207') CRSN='107'
IF (DEPTNUM=' 7' AND CRSNUM='231') CRSN='131'
```

Convert the equivalent departmental codes to departmental names.

```
IF (EDEPTNUM=' 1') EDEPT='EIL    '
IF (EDEPTNUM=' 2') EDEPT='BioChe '
IF (EDEPTNUM=' 3') EDEPT='Bio    '
IF (EDEPTNUM=' 4') EDEPT='Chem   '
IF (EDEPTNUM=' 5') EDEPT='Thca   '
IF (EDEPTNUM=' 6') EDEPT='Econ   '
```

Join the equivalent departments and course numbers to form the equivalent course, referred to as EQCRS.

```
COMPUTE EQCRS = CONCAT(EDEPT,CRSN)
```

Label the variables for clarity.

```
VARIABLE LABELS EDEPT 'DEPT NAME AFTER EQUIVALENCIES'
     CRSN 'COURSE NUMBER AFTER EQUIVALENCIES'
     EDEPTNUM 'DEPT CODE AFTER EQUIVALENCIES'
     EQCRS 'COURSE AFTER EQUIVALENCIES'
```

Save the system file and map the variables.

```
SAVE OUTFILE=CRS / MAP
```

These next commands refer to the raw assessment measure tape. In this example, the assessment measures are GRE scores. The raw data are from a tape that has one record per participant. The record length is 604.

```
DATA LIST FILE=IN3 /
          CENTER# 8-12
          TESTDATE 13-16
          SEX 41 (A)
          BRTHDATE 42-47
          BIQ1  109
          BIQ2  110-112
          BIQ3  113-114
          RACE 115
          BIQ5  116
          BIQ61 TO BIQ67  117-123 (A)
          BIQ7  124
          BIQ8  125
          BIQ9  126-127
          BIQ10 128-131
          BIQ11 132
          BIQ12 133-136
          BIQ13 137
          SEA1 TO SEA38  195-232 (A)
          SEB1 TO SEB30  245-274 (A)
          SEC1 TO SEC25  295-319 (A)
          SEE1 TO SEE38  345-382 (A)
          SEF1 TO SEF30  395-424 (A)
          SEG1 TO SEG25  445-469 (A)
          ID  600-603
FORMATS ID (F4.0)
```

The following recodes responses that were multiple codes to conform with the answer key provided. This particular tape had various codes to refer to a particular code. All codebooks will report this information if necessary.

```
RECODE SEA1 TO SEA38 ('A','B','C','D','*'='5') ('6','7','8','9'='0')
RECODE SEB16 TO SEB30 ('A','B','C','D','*'='5') ('6','7','8','9'='0')
RECODE SEC1 TO SEC25 ('A','B','C','D','*'='5') ('6','7','8','9'='0')
RECODE SEE1 TO SEE38 ('A','B','C','D','*'='5') ('6','7','8','9'='0')
RECODE SEF16 TO SEF30 ('A','B','C','D','*'='5') ('6','7','8','9'='0')
RECODE SEG1 TO SEG25 ('A','B','C','D','*'='5') ('6','7','8','9'='0')
RECODE SEB1 TO SEB15 ('6','A'='1') ('7','B'='2') ('8','C'='3')
                     ('9','D'='4') ('5','*'='0')
RECODE SEF1 TO SEF15 ('6','A'='1') ('7','B'='2') ('8','C'='3')
                     ('9','D'='4') ('5','*'='0')
```

The following converts string variables to numeric. This step may not be necessary if the assesment measures can be read as numeric. In the case of the GRE tape, as seen above, not all responses were coded as numeric.

```
RECODE SEA1 TO SEA38 (CONVERT) INTO SA1 TO SA38
RECODE SEB1 TO SEB30 (CONVERT) INTO SB1 TO SB30
RECODE SEC1 TO SEC25 (CONVERT) INTO SC1 TO SC25
RECODE SEE1 TO SEE38 (CONVERT) INTO SE1 TO SE38
RECODE SEF1 TO SEF30 (CONVERT) INTO SF1 TO SF30
RECODE SEG1 TO SEG25 (CONVERT) INTO SG1 TO SG25
```

Add variable labels.

```
VARIABLE
    TESTDATE 'TEST DATE'
    SEX  'SEX'
    BRTHDATE 'BIRTH DATE'
    ID  'STUDENT IDENTIFYING NUMBER'
    BIQ1  'CITIZENSHIP'
    BIQ2  'COUNTRY CODE'
    BIQ3  'STATE &TERRITORIAL CODE FOR RESIDENCE'
```

The following commands code the responses as either correct (1) or incorrect (0). This information should be provided with the assessment measures. The file created is GRE.SCORE.SPSS and will be used in the regression analysis.

```
RECODE SA1 (4=1) (ELSE=0)
RECODE SA2 (3=1) (ELSE=0)
RECODE SA3 (4=1) (ELSE=0)
RECODE SA4 (3=1) (ELSE=0)
RECODE SA5 (1=1) (ELSE=0)
RECODE SA6 (5=1) (ELSE=0)
RECODE SA7 (4=1) (ELSE=0)
RECODE SA8 (3=1) (ELSE=0)
RECODE SA9 (2=1) (ELSE=0)
RECODE SA10 (4=1) (ELSE=0)
```

THIS SECTION CREATES THE GRE ITEM-TYPES ACCORDING TO THE 1991 GRE KEY for one form of the test. You will need to modify according to the type of assessment instrument used and the procedures for calculating its scores.

```
COMPUTE  SC=SUM(SA1 TO SA7, SE1 TO SE7)
COMPUTE  ANA=SUM(SA8 TO SA16, SE8 TO SE16)
COMPUTE  RD=SUM(SA17 TO SA27, SE17 TO SE27)
COMPUTE  ANT=SUM(SA28 TO SA38, SE28 TO SE38)
COMPUTE  QC=SUM(SB1 TO SB15, SF1 TO SF15)
COMPUTE  RM=SUM(SB16 TO SB20,SB26 TO SB30,SF16 TO SF20,
    SF26 TO SF30)
COMPUTE  DI=SUM(SB21 TO SB25, SF21 TO SF25)
COMPUTE  AR=SUM(SC1 TO SC7,SC11 TO SC22,SG1 TO SG5,
    SG9 TO SG22)
COMPUTE  LR=SUM(SC8 TO SC10,SC23 TO SC25,SG6 TO SG8,
    SG23 TO SG25)
COMPUTE  VBRAWSC=SUM(SC,RD,ANT,ANA)
COMPUTE  QTRAWSC=SUM(DI,RM,QC)
COMPUTE  ANRAWSC=SUM(AR,LR)
COMPUTE  TORAWSC=SUM(VBRAWSC,QTRAWSC,ANRAWSC)
COMPUTE  TOFORMSC=SUM(VBRIGHTS,QTRIGHTS,ANRIGHTS)
COMPUTE  TOCONSC=SUM(VBCONVSC,QTCONVSC,ANCONVSC)

VARIABLE LABELS VBRAWSC 'SUM OF RAW VERBAL SCORES'
 SC 'SENTENCE COMPLETION'
 ANA 'ANALOGY'
 RD 'READING COMPREHENSION'
 ANT 'ANTONYMS'
 QC 'QUANTITATIVE COMPARISON'
 RM 'DISCRETE QUANTITATIVE'
 DI 'DATA INTERPRETATION'
 AR 'ANALYTICAL REASONING'
 LR 'LOGICAL REASONING'
```

In order to match system files of this type, is is necessary to have all data sets in the same order according to some variable. In this case, the variable is the ID that was given to each participant. This will insure that each individual is matched properly. The command below sorts cases in an ascending sort.

```
SORT CASES BY ID
```

Save the file and drop unneeded converted variables.

```
SAVE OUTFILE=GRE / DROP SEA1 TO SEA38, SEB1 TO SEB30,
      SEC1 TO SEC25,SEE1 TO SEE38, SEF1 TO SEF30,
      SEG1 TO SEG25 / MAP
GET FILE=DEMO / KEEP ID SATM SATV SATT
```

<u>Please Note</u>: you will need to match files to include sat scores on the gre file to run regression, assumes demo system file is already made.

```
SORT CASES BY ID
```

The two files will be joined by matching ID's. The second file in this command is represented by an asterik to denote the active file.

```
MATCH FILES FILE=GRE/FILE=*/BY ID
```

The following commands perform the regression analysis used in the DCP. The command /SAVE RESID(variable name) saves the residuals to the active file and names the residuals 'variable name'.

```
REGRESSION VARS SATV SC ANA RD ANT
 /DEP=SC
 /ENTER
 /SAVE RESID(RSC)
 /DEP=ANA
 /ENTER
 /SAVE RESID(RANA)
 /DEP=RD
 /ENTER
 /SAVE RESID(RRD)
 /DEP=ANT
 /ENTER
 /SAVE RESID(RANT)

REGRESSION VARS SATM QC RM DI
 /DEP=QC
 /ENTER
 /SAVE RESID(RQC)
 /DEP=RM
 /ENTER
 /SAVE RESID(RRM)
 /DEP=DI
 /ENTER
 /SAVE RESID(RDI)
REGRESSION VARS SATT AR LR
 /DEP=AR
 /ENTER
 /SAVE RESID(RAR)
 /DEP=LR
 /ENTER
 /SAVE RESID(RLR)
```

Note: To reduce the amount of space used in analysis and for convenience in reanalysis the residuals and other relevant variables are saved in a separate file called

GRE.RESID.SPSS

```
SAVE OUTFILE=RES / KEEP ID SC ANA RD ANT QC RM DI AR
   LR RSC RANA RRD RANT RQC RRM RDI RAR RLR SATV
   SATM SATT / MAP
GET FILE=CRS / KEEP ID EQCRS COURSE
SORT CASES BY ID
```

The next set of commands match the residual file with the course file. It is necessary to use the TABLE command to indicate that all courses with identical ID's should be matched with the ID's residual.

```
MATCH FILES TABLE=RES/FILE=*/BY ID
XSAVE OUTFILE=CRS1 / MAP
```

The following commands compute the mean residuals for each course and save the results to the active file. The ENROLL=N command puts the number of times a course appeared on the transcripts into the variable ENROLL.

```
AGGREGATE OUTFILE=*
 /BREAK=EQCRS
 /ENROLL=N
 /MRSC MRANA MRRD MRANT MRQC MRRM MRDI MRAR MRLR=
  MEAN(RSC RANA RRD RANT RQC RRM RDI RAR RLR)
VARIABLE LABELS ENROLL 'NUMBER OF STUDENTS TAKING
COURSE'
  MRSC 'MEAN RESIDUAL SC'
  MRANA 'MEAN RESIDUAL ANA'
  MRRD 'MEAN RESIDUAL RD'
  MRANT 'MEAN RESIDUAL ANT'
  MRQC 'MEAN RESIDUAL QC'
  MRRM 'MEAN RESIDUAL RM'
  MRDI 'MEAN RESIDUAL DI'
  MRAR 'MEAN RESIDUAL AR'
  MRLR 'MEAN RESIDUAL LR'
```

Save the relevant variables to a new file.

```
XSAVE OUTFILE=MRES / KEEP MRSC MRANA MRRD MRANT
      MRQC MRRM MRDI MRAR MRLR EQCRS / MAP
```

In the DCP Project, a criteria of an enrollment of 5 students was used to select the courses to be analyzed.

```
SELECT IF (ENROLL > 4)
```

The following commands calculate the similarity matrix. The matrix then becomes the active file due to the command MATRIX=OUT(*).

```
PROXIMITIES MRSC MRRD MRANA MRANT MRQC MRRM MRDI
        MRAR MRLR
/VIEW=CASE
/MEASURE=CORRELATION
/ID=EQCRS
/MATRIX=OUT(*)
/PRINT=NONE
```

The following job control language performs the cluster analysis. The analysis will use the active file created in proximities. The PRINT=SCHEDULE(5,25) determines the number of different cluster groupings that will be printed. It is necessary to determine this amount as it is this schedule printout will be the data used in the discriminant analysis. In this example, the range of of clusters printed is from the 5-cluster solution to the 25-cluster solution as indicated by the numbers in parentheses in the PRINT command.

```
CLUSTER
/MATRIX=IN(*)
/ID=EQCRS
/METHOD=BAVERAGE
/PRINT=SCHEDULE CLUSTER(5,25)
/PLOT=DENDROGRAM
```

Up to this point, the analysis should run with only a change in the job control language to match your facilities. The spssx version that we currently have will no longer save the clusters to a file that has been created by a matrix input. Your issue may. if not, you must create a system file from the output of the cluster procedure via editing the output to create a raw data base that has the courses and their respective cluster membership (this information is found near the end of the printout, right before the dendogram in the Schedule. Insert the data into the systemfile creating program (inline) below with the appropriate columns designated and the appropriate job control language to create the new system file that will contain the cluster data. The program that follows the making of the cluster system file will then match the Cluster file to the Mres file made previously. Once this is accomplished, the following discriminant analysis can be run.

The following job control language has been included to show files that are necessary.

```
//V03XXXXX JOB
// EXEC SPSSX,PARM=250K
//CLUS                                                              DD
DSN=MEN.U69640.HTS.QUANCLUS.SPSS,DISP=(NEW,KEEP),
// VOL=REF=MEN.U69640.HTS.LIB,SPACE=(TRK,(20,10),RLSE)
//MRES                                                              DD
DSN=MEN.U69640.HTS.CRS.MRESID.SPSS,DISP=(OLD,KEEP),
// VOL=REF=MEN.U69640.HTS.LIB
//QUAL                                                              DD
DSN=MEN.U69640.HTS.QUAN.CLUS.SPSS,DISP=(NEW,KEEP),
// VOL=REF=MEN.U69640.HTS.LIB,SPACE=(TRK,(20,10),RLSE)
//FILE HANDLE CLUS
DATA LIST FILE=INLINE FIXED
 1 COURS 1-10 (A) OBS 14-16 CLUSM25 22-23 CLUSM15 28-29
   CLUSM13 33-34 CLUSM11 38-39 CLUSM10 43-44 CLUSM9 49
   CLUSM8 54 CLUSM7 59 CLUSM6 64 CLUSM5 69
VALUE LABELS CLUSM13 1 '1ST CLUSTER MEMBER'
 2 '2ND CLUSTER MEMBER'
 3 '3RD CLUSTER MEMBER'
 4 '4TH CLUSTER MEMBER'
 5 '5TH CLUSTER MEMBER'
 6 '6TH CLUSTER MEMBER'
 7 '7TH CLUSTER MEMBER'
 8 '8TH CLUSTER MEMBER'
 9 '9TH CLUSTER MEMBER'
 10 '10TH CLUSTER MEMBER'
 11 '11TH CLUSTER MEMBER'
 12 '12TH CLUSTER MEMBER'
 13 '13TH CLUSTER MEMBER'
 BEGIN DATA
```

INSERT INLINE DATA HERE AS DEFINED BY YOUR INLINE DATA
STATEMENT ABOVE

END DATA
SAVE OUTFILE=CLUS

Following is an example of the portion of the cluster printout you will want to use to make the inline data.

CLUSTER MEMBERSHIP OF CASES USING AVERAGE LINKAGE (BETWEEN GROUPS)

NUMBER OF CLUSTERS

LABEL	CASE	15	14	13	12	11	10	9	8	7	6	5
Anthro 103	1	1	1	1	1	1	1	1	1	1	1	1
Anthro 104	2	2	2	2	2	2	2	2	2	2	2	2
Anthro 105	3	1	1	1	1	1	1	1	1	1	1	1
Anthro 129	4	2	2	2	2	2	2	2	2	2	2	2
Anthro 270	5	1	1	1	1	1	1	1	1	1	1	1
Anthro 280	6	1	1	1	1	1	1	1	1	1	1	1
Anthro 290	7	3	3	3	3	3	3	3	3	3	3	3
Anthro 293	8	2	2	2	2	2	2	2	2	2	2	2
Anthro 302	9	1	1	1	1	1	1	1	1	1	1	1
Anthro 320	10	1	1	1	1	1	1	1	1	1	1	1

The following commands run the discriminant analysis. It is necessary to match the cluster file just created to the previously created Mres file to obtain the course mean residuals. The example shows the cluster analysis of a 13 cluster solution.

```
GET FILE=MRES
SELECT IF (ENROLL > 4)
MATCH FILES FILE=CLUS/FILE=*

TITLE 'DISCRIMINANT ANALYSIS OF 13-CLUSTER MEMBER'
SUBTITLE 'ENROLL > 4 USING DIRECT DISCRIMINATING
METHOD'
SUBTITLE 'USING QUANTITATIVE CLUSTER GROUPING'
DISCRIMINANT GROUPS=CLUSM13(1,13)/
 VARIABLES=MRANA MRANT MRSC MRRD MRQC MRRM MRDI
MRAR MRLR/
 METHOD=DIRECT/
 PRIORS=SIZE/
 STATISTICS=RAW COEFF TABLE/
 PLOT=
```

REFERENCES

Adelman, Clifford. (1988). Linking student outcomes and the curricular reform: A new research model. A forum presented at the annual meeting of the American Association for Higher Education, Washington, DC.

American Association of Colleges. (1985). Integrity in the college curriculum: A report to the academic community. Washington, DC: Association of American Colleges.

Association of American Colleges. (1988). A new vitality in general education. Washington, DC: Association of American Colleges.

American Council on Education. (1991). Assessing assessment. Washington, DC: Author.

Astin, A. W. (1970a). The methodology of research on college impact, part 1. Sociology of Education, 43, 223-254.

Astin, A. W. (1970b). The methodology of research on college impact, part 2. Sociology of Education, 43, 437-450.

Beeken, L. A. (1982). The general education component of the curriculum through transcript analysis at three Virginia community colleges. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University.

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

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

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

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.

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

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 distribution: Amount and specialization. Journal of Higher Education, 58, 430-442.

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

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

Eaton, J. S. (1990). Three myths of transfer education. Community, Technical and Junior College Journal, 60, 18-20.

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

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

Gardiner, J. et al. (1986). Instructional resources in higher education, 2nd edition. Stillwater, OK: Oklahoma State University.

Jones, Elizabeth A., & Ratcliff, James L. (1990a). Is a core curriculum best for everybody? The effect of different patterns of coursework on the general education of high and low ability students. Paper presented at the annual meeting of the American Educational Research Association, Boston, Massachusetts.

Jones, Elizabeth A., & Ratcliff, James L. (1990b). Effective coursework patterns and faculty perceptions of the development of general learned abilities. Paper presented at the annual meeting of the Association for the Study of Higher Education, Portland, Oregon.

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

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

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), 281-288.

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

Pascarella, E. T., & Terenzini, P. T. (1991). How colleges affects students: Findings and insights from twenty years of research. San Francisco, CA: Jossey-Bass Publishers.

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, NY: Holt, Rinehart, and Winston.

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

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

Ratcliff, James 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. Ames, IA: Iowa State University.

Ratcliff, James L. (1988a). The development of a cluster analytic model for determining the associated effects of coursework patterns on student learning. A paper presented at the annual meeting of the American Education Research Association (AERA), New Orleans, Louisiana.

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

Ratcliff, James L. (1990a). Development and testing of a cluster-analytic model for identifiying coursework patterns associated with general learned abilities of college students (Final Report, Stanford University Samples #1 and #2). U.S. Department of Education, Office of Educational Research and Improvement, Contract No. OERI-R-86-0016. Ames, IA: Iowa State University.

Ratcliff, James L. (1990b). Development and testing of a cluster-analytic model for identifiying coursework patterns associated with general learned abilities of college students (Final Report, Ithaca College Samples #1 and #2). U.S. Department of Education, Office of Educational Research and Improvement, Contract No. OERI-R-86-0016. Ames, IA: Iowa State University.

Ratcliff, James L. (1990c). Development and testing of a cluster-analytic model for identifiying coursework patterns associated with general learned abilities of college students (Final Report, Ithaca College Sample #3). Exxon Education Foundation. University Park, PA: Center for the Study of Higher Education, The Pennsylvania State University.

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

Ratcliff, James L., & Jones, Elizabeth A. (1990). General learning at a women's college. Paper presented at the annual meeting of the Association for the Study of Higher Education, Portland, Oregon.

Ratcliff, James L., & Jones, Elizabeth A. (1991). Are common course numbering and a core curriculum valid indicators in the articulation of general education credits among transfer students? Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois.

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).

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

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

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

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

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

Tyler, R. (1950). Basic principles of curriculum development. Chicago, IL: 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, NY: McGraw-Hill.

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