DOCUMENT RESUME

ED 293 379 HE 020 047

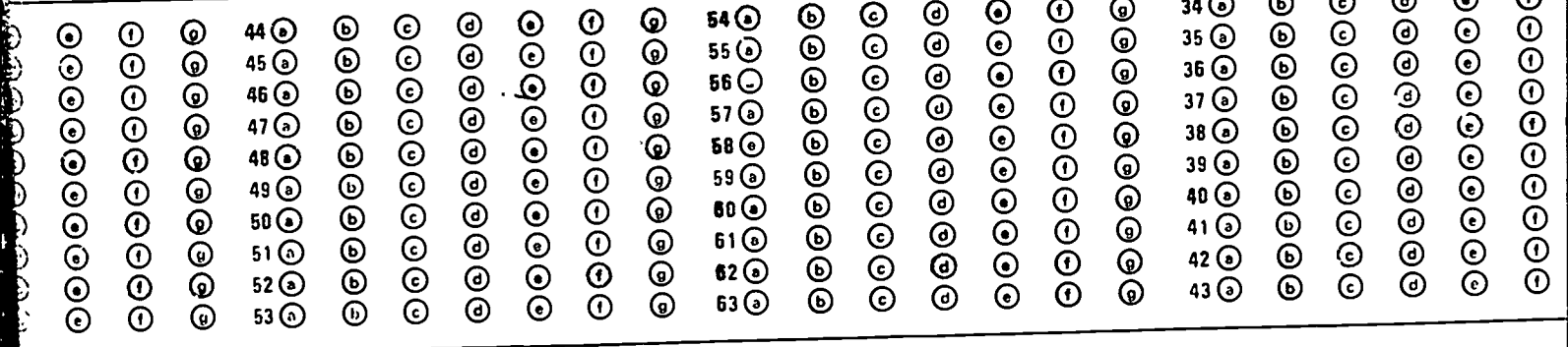
TITLE	Defining and Assessing Baccalaureate Skills: Ten Case Studies. A Report on the Academic Program Evaluation Project.
INSTITUTION	American Association of State Colleges and Universities, Washington, D.C.
SPONS AGENCY	Fund for the Improvement of Postsecondary Education (ED), Washington, DC.
REPORT NO	ISBN-0-88044-065-1
PUB DATE	86
NOTE	81p.; Paper collected as part of the American Association for Higher Education Assessment Forum. Case study institutions also include Western Kentucky University.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Bachelors Degrees; Case Studies; *College Curriculum; Evaluation Criteria; *General Education; Higher Education; *Outcomes of Education; *Program Evaluation; Skill Development; State Colleges; State Universities; *Student Evaluation
IDENTIFIERS	*AAHE Assessment Forum; Ball State University IN; California State University Chico; College Outcomes Assessment; North Adams State College MA; Ramapo College of New Jersey; Southern Illinois University Edwardsville; State University of New York Coll at Potsdam; University of Nebraska Omaha; Wayne State College NE; Western Carolina University NC; Western Kentucky University

ABSTRACT

Efforts to evaluate the general education component of baccalaureate degree programs at ten colleges using a uniform assessment formula are described. The colleges were participants in the American Association of State Colleges and Universities' Academic Program Evaluation Project (APEP), which proposed that outcomes of baccalaureate education can be expressed as generic intellectual skills that are measurable and by which students can be evaluated. Information is provided on the five stages of APEP: defining generic skills; identifying performance indicators, criteria, and testing procedures; assessing students and programs; judging student and program performance; and policy analysis and decision making. Institutional profiles, background information, and descriptions of the evaluation efforts in the early 1980s are provided for following colleges and universities: Ball State University; California State University, Chico; North Adams State College; Ramapo College of New Jersey; Southern Illinois University at Edwardsville; State University College at Potsdam in New York; University of Nebraska at Omaha; Wayne State College; Western Carolina University; and Western Kentucky University. (SW)

A Report
on the
Academic Program
Evaluation Project

Defining and Assessing Baccalaureate Skills



Ten Case Studies

American Association
of State Colleges and Universities

A Report
on the
Academic Program
Evaluation Project

Defining and Assessing Baccalaureate Skills

Ten Case Studies

Produced with major support
from the Fund for the Improvement
of Postsecondary Education

American Association
of State Colleges and Universities

One Dupont Circle
Suite 600
Washington, D.C. 20036
202/293-6440

AMERICAN ASSOCIATION
FOR HIGHER EDUCATION

The AAHE ASSESSMENT FORUM is a three-year project supported by the Fund for the Improvement of Postsecondary Education. It entails three distinct but overlapping activities:

--an annual conference
(the first scheduled for June 14-17, 1987, in Denver)

--commissioned papers
(focused on implementation and other timely assessment concerns; available through the Forum for a small fee)

--information services
(including consultation, referrals, a national directory, and more)

This paper is part of an on-going assessment collection maintained by the Forum. We are pleased to make it more widely available through the ERIC system.

For further information about ASSESSMENT FORUM activities, contact Patricia Hutchings, Director, AAHE ASSESSMENT FORUM, One Dupont Circle, Suite 600, Washington, DC 20036

Library of Congress Cataloging-in-Publication Data

Defining and assessing baccalaureate skills.

1. Universities and colleges--United States--Curricula--Evaluation--Case studies. 2. Bachelor of arts degree--United States--Evaluation--Case studies. 3. Bachelor of liberal studies--United States--Evaluation--Case studies. I. American Association of State Colleges and Universities.
LB2361.3.D44 1986 378'.199'0973 86-7963
ISBN 0-88044-065-1

Contents

Preface

Few of us in higher education would dispute the notion that college graduates should be able to demonstrate readily that they possess certain skills. Equally acceptable is the notion that among those skills must be the ability to communicate effectively, to engage in analytical and synthetical reasoning, to compute accurately and to solve quantitative problems of moderate complexity, and perhaps to develop a personal value system. Consensus begins to disappear, however, on definitions of each of these skills. In addition, if one further asks the classical who? what? how? where? and when? the answers are not only elusive, but widely varying. The reports presented here describe the efforts of ten campuses to find answers.

These efforts arose from a project largely underwritten by a grant from the Fund for the Improvement of Postsecondary Education. Eventually known as the Academic Program Evaluation Project (APEP), the project was based on the assumption that a modular approach could in fact facilitate defining and answering the questions about academic skills. The project began with ten campuses serving as experimental sites at which a multistep model was to be tested. Under the direction of Marina Buhler-Miko, then the director of AASCU's Resource Center for Planned Change, teams of individuals from the ten participating institutions began to develop campus-specific approaches toward the ultimate objectives. There was a strong consensus among the project participants that every campus had unique circumstances calling for individualized approaches.

At the beginning of the project, the multistep process was available in limited detail. Later, it became clear that much more detail and an expanded theoretical base for each step were needed. Those needs should become apparent as one reads the reports from the ten campuses. Work in this regard has been proceeding through efforts of Robert Stakenas and Gary Peterson of Florida State University in collaboration with Ms. Buhler-Miko, who is now president of the Higher Education Strategic Planning Institute.

By disseminating the experiences of the ten APEP campuses, AASCU hopes to foster awareness of the necessity to give serious attention to the processes by which collegiate generic skills are to be developed; of the wide diversity of points of view that prevail on a single campus; of the need for efforts directed toward reaching consensus on how each campus defines each skill and its satisfactory development; finally, of the view that rigid national standards are next to impossible given the variety of institutions in this country. At the same time, every college and university must have a plan and program that will lead to demonstrable effectiveness in developing student skills as understood by each institution.

Harold Delaney
Executive Vice President
American Association of State
Colleges and Universities

Introduction

This compendium of case studies describes the efforts in the early 1980s of ten members of the American Association of State Colleges and Universities (AASCU) to evaluate the general education component of their baccalaureate degree programs by means of a uniform assessment formula.

The participants in AASCU's Academic Program Evaluation Project (APEP) were ahead of their time. Since their institutional evaluations took place, U.S. higher education has come under sharp attack, both from within and outside the profession, for allegedly fashioning curricula that are too utilitarian, failing to involve either students or faculty members systematically in "the learning process," and neglecting or ignoring the liberal arts. General education has suffered, the critics contend. Specialization dominates the undergraduate program, with lack of an institutionwide "sense of mission" the tragic result.

Clearly, the latter half of the 1980s will be a most difficult period for American colleges and universities. In response to the critics, many institutions will undergo some form of self-analysis. For a few, the experience--unhappily--may be too late to be helpful. For the more fortunate, it may occur--as it did for the colleges and universities involved in APEP--at a time conducive to campuswide change, resulting in enhancement of the learning process for both students and faculty members, enrichment of the curriculum, strengthening of general education, and renewed sense of institutional mission.

In one of the critiques of undergraduate programs published in late 1984, the Study Group on the Conditions of Excellence in American Higher Education, sponsored by the National Institute of Education, reported that "the quality of undergraduate education could be significantly improved if America's colleges and universities would apply existing knowledge about three critical conditions of excellence--(1) student involvement, (2) high expectations, and (3) assessment and feedback."

The study group was voicing precisely the concerns that motivated AASCU, in 1979, to initiate the Academic Program Evaluation Project (APEP), with partial funding by the Fund for the Improvement of Postsecondary Education (FIPSE). Many AASCU members perceived a need for evaluation of their baccalaureate degree programs, and APEP, as conceived, affirmed that there should be certain generic, recognizable outcomes of every such program. The project's objective was to establish a process by which all constituents of a college or university might specify the intellectual skills and abilities expected to accompany the earning of a bachelor's degree, prescribed in a manner acceptable to the constituents and in accordance with the aims of the particular institution. The process would necessarily include techniques by which the outcomes could be evaluated.

APEP's focus was to be broad rather than narrow. As an outgrowth of discussions on the meaning of the baccalaureate, the assessment of the quality of intellectual skills required for a bachelor's degree, and the reform of general education, the project was designed to provide means for evaluation of an institution's educational results. The descriptive term *generic* became useful for highlighting those intellectual skills and abilities which would be instructed, promoted, and required in all the disciplines, and which therefore might be said to characterize the recipient of a bachelor's degree, regardless of major or type.

Recognizing the need for a variety of plans from which other institutions could adopt those features applicable to their own circumstances, ten institutions were awarded FIPSE grants to implement the project on their respective campuses. These were Ball State University (Indiana); California State University--Chico; North Adams State College (Massachusetts); State University College of Arts and Science--Potsdam (New York); Ramapo College of New Jersey; Southern Illinois University at Edwardsville; University of Nebraska at Omaha; Wayne State College (Nebraska); Western Carolina University (North Carolina); and Western Kentucky University.

Under guidance of AASCU's Resource Center for Planned Change, the ten participating institutions formed committees of administrators and faculty members. While each institution would conduct its work as if the project were unique to its own campus, the committees formed a network for information and experience exchanges in conjunction with consultants and the staff of the center. The first consultants, J. R. Warren of the Educational Testing Service (ETS) and Robert Stakenas of Florida State University, were joined later by Eldon Park of ETS, Gary Peterson of

Florida State, and Gary Woditsch of Bowling Green University (Ohio).

The Project: Outcomes of the Baccalaureate

APEP was based on the assumption that the outcomes of baccalaureate education can be expressed as generic intellectual skills; that these skills are the common denominators of the undergraduate curriculum; that they can be defined and broken down into measurable components; and that the extent to which they have been acquired by students can be evaluated. It followed, then, that if generic skills could be defined sufficiently to be measured, they could be used to evaluate the effectiveness of academic programs and, thus, of educational institutions. For the purposes of the project, a *generic skill* was defined as a complex set of intellectual abilities that can be generalized and are applicable to all the academic disciplines, and, ideally, would therefore be demonstrated by all graduating students.

The idea of evaluating undergraduate education in relation to transinstitutional, generic outcomes developed over many decades, but special attention was focused on actually implementing such an evaluation when, in 1978, the American Council on Education's Task Force on Educational Credit and Credentials reported the following recommendations:

> Associate degrees and Bachelor's degrees should attest to at least the following three types of accomplishment:
>
> 1. Accomplishment specified by the awarding institution as necessary for the development of a broadly educated person, including familiarity with general areas of knowledge;
> 2. Competence in analytical, communication, quantitative; and synthesizing skills;
> 3. Accomplishment in a specialized area of study covering a set of integrated learnings requiring analysis, understanding or principals that have judgmental application, and a theoretical knowledge base.
>
> Each credential-granting institution should clearly define, to the extent possible, the meaning of the certificates and degrees it awards.

Following up on these recommendations, J. R. Warren of ETS undertook research to discover whether it was possible to measure such "generic" outcomes as analysis, communication, quantification, and synthesis. At the same time, the Center Associates of AASCU's Resource Center for Planned Change agreed that the same four generic skills should become the common base, yielding individualized interpretations for the ten APEP institutions. A fifth skill--valuing--was also identified and could be included at the option of each participating institution.

It should be noted that APEP was not the only project expressing educational outcomes as generic skills. Research reports in educational psychology had suggested that intellectual skills traverse subjects and disciplines, that learning of skills proceeds in a hierarchical fashion, and, thus, that individuals have varying levels of skill development that are independent of subject-matter mastery. These fundamental developments in learning theory brought on the "interdisciplinary" and then the "competency-based" curricular trends of the 1970s. During this time, the College Board began its ten-year Equality Project to address preparation for college by defining the learning skills students should have in order to succeed in college. School and college faculty members and other educators nationwide participated in identifying six basic academic competencies: reading, writing, speaking and listening, mathematics, reasoning, studying. In a second phase of the project, six basic academic subjects were identified for secondary education: English, the arts, mathematics, science, social studies, foreign language. More recently (1984), AASCU published *In Pursuit of Degrees with Integrity*, describing "a value-added approach to undergraduate assessment" adopted by Northeast Missouri State University.

By using generic skills to define the fundamental outcomes of baccalaureate education, it was possible to design an evaluation process that would adjust for different institutional goals and involve faculty members from all disciplines. (The five-stage evaluation process used by the APEP participating institutions is outlined in the next section of the Introduction.) Once the ten institutions had identified the generic intellectual skills their respective general education programs sought to produce in their students, the evaluation process consisted of defining each skill in measurable terms, selecting instruments and procedures for measurement, assessing student acquisition of generic skills, and evaluating the results in relation to the academic program. Policy implications emerged as the evaluation proceeded at each institution.

Because the APEP institutions had agreed on communication, quantification, analysis, synthesis, and valuing as the relevant generic skills for the project, at each campus the work began with attempts to define these skills in specific, measurable terms. The campus project directors and the interdisciplinary teams or committees proceeded to construct "conceptual

definition, component subskills, observable behaviors and levels of performance" for each generic skill. They were assisted in this task by the project consultants. Although AASCU's Resource Center staff provided guidelines for the process of defining generic skills, each campus approached the task in its own fashion, as the ten case studies prepared by the campus project directors reflect. Similarly, each institution followed its own procedures on other phases of the project, from the composition of committees to the method of decision making.

The Five Stages of APEP: The Evaluation Process

The five stages of the Academic Program Evaluation Project were designed to direct the attention of faculty members and administrators to key issues in the assessment of generic skills. The stages were intended to provide guidance, not a prescription, and the scope was deliberately planned to be broad enough to accommodate diverse institutional missions, curricula, instructional styles, standards, educational policies, planning, and measurements.

Faculty and administrative representatives of institutions participating in the project were expected to work together to

- identify institutional aims in terms of generic intellectual skills
- define each of the generic skills valued by the institution in measurable terms
- determine where and how well those skills were being taught
- pinpoint curricular and instructional instances in which generic skill performance was below required standards
- recommend policy and procedural changes designed to raise performance levels in any areas of low generic skills.

Procedurally, these basic tasks were conceived as a developmental process, and thus developmental terms seemed most appropriate in suggesting the steps through which the participating institutions would pass. The APEP process is, therefore, most easily described in terms of "stages," with certain tasks to be accomplished during each stage.

Stage I: Developing Definitions of Generic Skills

An institution that seeks to evaluate its academic program typically begins by defining the generic skills it seeks to encourage in its students. Although APEP focuses on the specific skills of analysis, synthesis, quantification, communication and valuing, the skills chosen could be any generic skills valued by the faculty. Formulated in concise, conceptual terms, these skill definitions are the guides for all subsequent assessment steps.

By definition, generic intellectual skills are cross-disciplinary and were considered, at least initially, in discipline-free terms. They reflect the essential and common characteristics of thinking processes that occur in every field of study. Once the abstract statements of capabilities are agreed on, the work then proceeds to the identification of subcomponents or subskills that can demonstrate the capabilities. After these definitions are submitted for faculty consideration and approval, Stage II begins.

Component Tasks

1. Determine the purpose of an evaluation project, whether as broad as the entire baccalaureate level or as specific as selected professional programs. Establish the intended audience for project results. Determine the project scope in order to identify the period of time (e.g., freshman and sophomore years), groups of students and programs, and any special attributes.

2. Review the institutional mission to identify the generic skills inherent in the stated purposes of the college or university.

3. Prepare conceptual definitions of the generic skills and subskills that are measurable. Gradations in performance levels need to be described. Distribute these definitions and associated features to the faculty for critical response and ultimate acceptance.

4. From among the defined generic skills and subskills, select those for inclusion in the evaluation. (Ordinarily, it would be too ambitious and predictably unwise to pursue measurement of all the skills.)

Stage II: Identifying Performance Indicators, Criteria, and Testing Procedures

The definitions of generic skills become useful only if observable indicators are discovered. Such indicators should be sought in the intellectual behaviors demanded of students, indicating various levels of competence in various disciplines. Generic skills cannot be observed directly, and these "indicator" behaviors or performances permit evaluators to draw inferences about the skills.

The major task becomes one of identifying and developing testing procedures to evoke student performances indicative of generic skill development. Statements describing levels of performance--medium, high, low--must be formed and criteria established to judge the level of student development. Because generic skills are multifaceted, it is unlikely that

they could be adequately assessed by a single procedure. Multiple performance measures with varied complexity and subject-matter content of the assignment present students with opportunities to demonstrate their competencies in differing intellectual situations.

Component Tasks

1. On the basis of the project mission, select and validate the scope, skills, and content of the project. "Scope" denotes the context in which the skills are developed (e.g., a course, a major, a baccalaureate degree, life in general). "Skill" is the generic skill to be assessed. "Content" refers to the complexity and breadth of the domain wherein the skill development is believed to occur (e.g., a topic, a discipline, a cluster of fields, an entire program).

2. Design the assessment procedures. Direct measures of performance include all forms of written, oral, and artistic works whereby the student's product or his contribution might be assessed. Indirect measures can be used to evaluate a student's choice of responses that are indicative of future performance.

Beyond selecting the assessment procedures, take steps to ensure the availability of materials, resources, and guidance necessary to prevent irrelevant handicaps to student performances.

3. Develop specifications for tests and test items. The preparation of locally developed tests calls for available professional talent to guide and validate the product. This effort requires great support from the faculty. Pilot test all instruments purchased or devised.

4. Establish a general policy and process for all tests, whether home developed or purchased commercially. The process should provide for a general analysis of the merits and liabilities of locally produced *vs.* commercially available tests.

Stage III: Assessment of Students and Programs

There are two tasks for this stage: assessing of generic skill development among students, and determining components of an academic program in which a student has the opportunity to acquire or improve generic skills.

Because it is not feasible to measure every student's development and not possible to observe every learning experience, decisions are needed on which students to sample, how many students to sample, and the extent of academic programs experienced.

A plan for data collection, analysis, and interpretation must be drawn up.

Component Tasks

1. Formulate the learning outcome evaluation design--i.e., the framework for decisions about where, when, and from whom data will be collected. Recognize that a control group and rigorous sampling procedures would not be possible initially.

2. Specify the instructional events that develop generic skills; that is, in what parts of the instructional program generic skills are likely to be developed.

3. Identify sequences of courses or other identifiable learning experiences and endeavor to learn which of the skills is likely to be affected by the elements of sequences. Once the identity is established, a sampling procedure can be found and data collection begun.

Stage IV: Evaluation--Judging Student and Program Performance

The data accumulated in the assessment of student progress must be interpreted against standards presumed relevant to the institutional mission and student body. Traditionally, the standards are those fixed by the faculty.

Student performance standards can be relative--i.e., compared with students in other places or with students of a prior year. Standards can also be absolute, as defined in faculty convictions about the goals of a baccalaureate.

Comparing performance data with the standards establishes whether the standards have been met; whether the intents and the observations correspond; and whether the gaps discerned in student performance are serious enough to warrant action.

Component Tasks

1. Arrange all data accumulated relative to programs and students for analysis. It is often convenient to divide the information into antecedents (givens), transactions, and outcomes in order to analyze and describe of what was intended and what was observed.

2. Determine the congruence between what was intended and that which was observed. Lack of congruence could indicate that the learning assignments did not effect the desired skills, or that the applied measurements were not adequate to assess the skill in question.

3. Determine the logical and empirical contingencies. The former are based on prior experience and show reason in the planning of instruction and assignments and are conguent with the course objectives; tests at the end should also be congruent. Empirical contingencies require a kind of audit of

faculty members and students, done independently to judge the performance of both--ultimately, to discover any lack of congruence in skill development.

4. Compile a list of discrepancies and performance gaps.

5. Determine standards for making judgments. A comprehensive program assessment would judge more than student outcomes; it would focus as well on the "givens" and the entire intellectual strategy. Hence the standards applied to the whole, including the levels of expectation of students, becomes the heart of the assessment process. Setting standards is an act of judgment, and participants should exercise great care in order to ensure fairness and responsibility.

6. On the basis of standards, make a judgment of worth--that is, a comparison of what was observed and the prescribed standards, with discrepancies noted and reviewed for levels of importance. Those deemed critical are reserved for consideration in Stage V.

Stage V: Policy Analysis and Decision Making

If discrepancies were noted in Stage IV, determine the causes. Did the discrepancies arise because of limited resources or inappropriate use of resources? Could the cause be traced to admissions, curricula, instructional practices, or some other variable? If the causes are found, remedies can be devised and assessed for potential effectiveness and feasibility.

In this stage, consider alternatives for the remedy or the neglect of discrepancies. A decision to apply a remedy could cause the institution to revisit another stage. For example, if a decision is made to redraw original definitions, Stage I would be reentered. If different testing procedures is indicated, Stage II would be reentered. If the definitions and testing procedures remain the same, then learning outcomes and program effectiveness could be reviewed in Stage III.

Component Tasks

1. Determine the reasons for performance gaps in observed outcomes. In some instances, the component steps outlined in Stage IV could assist in this determination.

2. Compare the cost of removing the gaps with the consequences of ignoring the gaps.

3. Arrange the gaps in learning outcomes according to importance and select gaps on which actions are to be taken.

4. Determine policy changes required for removal of the gaps. Revised goal statements relative to the generic skills are often necessary. Sometimes academic goals require revision to clarify expectations about the skills.

5. Make changes in instructional practices necessary to remove gaps. Significant changes in student and faculty role behavior are frequently needed.

6. Encourage participation of the faculty in making decisions about policy and instructional changes. This step has proven crucial to success.

7. Devise a time schedule for completion in order to facilitate evaluation of the effectiveness of the changes.

Following are descriptions of the APEP experiences of ten AASCU institutions, together with a summary of the policy and administrative issues raised by academic program evaluation on the participating campuses.

Ball State University (IN)

Institutional Profile.

Ball State University is a comprehensive, publicly assisted institution of higher education that aspires to excellence in teaching, research and other scholarly productivity, and public service. The university's missions and constituencies are regional, statewide, and national in character, depending on the program or activity. The university also engages in a selected number of programs distinctly international by virtue of their location and clientele.

Ball State University is an academic community substantially residential in character but nonetheless one in which outreach activities, continuing education, and the provision of education in a variety of off-campus locations loom large. Moreover, the university is dedicated to the pursuit of truth and the provision of a liberal education to all its students regardless of their majors or locations.

The university offers a comprehensive range of academic programs at the associate, baccalaureate, and master's levels, as well as doctoral programs in areas in which it has special competence and the State of Indiana has special needs. In addition to the core academic programs in the arts, sciences, and humanities that every comprehensive university provides, Ball State also offers a broad range of professional programs that not only prepare students for specific occupational and vocational interests but also provide them with a liberal education. Important among these professional offerings is a complete and diverse range of teacher education programs that build upon the university's traditional strengths. These professional offerings also include a widening circle of programs in such areas as actuarial science, allied health sciences, architecture, landscape architecture, urban and regional planning, nursing, bioenergetics, a wide range of business and management-related disciplines, computer science, library science, public administration, social work, and speech pathology. Ball State University is committed to furthering professionally oriented, programmatic growth in those areas wherein it exhibits competence and there is demonstrated need.

The emphasis of the university's major academic programs during the next decade will be concentrated in areas that support the provision of human services; the preparation of liberally educated students in professional and vocational fields; and the solution of societal problems, including those pertaining to energy, high technology, resource management, health, and the quality of life. Some of these programs will be regionally oriented by virtue of the university's special relationship with its geographic area. Others, such as architecture and planning, teacher education, and the bioenergetics graduate program, will reflect not only state needs and constituencies but national ones as well.

The university's policies governing admission and retention of students are based on a preference for students who exhibit either immediate ability or strong potential to benefit from higher education at Ball State. Hence, the university is selective and does not accept for admission all students who wish to enter. In certain areas, the university's admissions policy is highly selective. The university's interest in students of high ability and potential transcends variables such as age, handicap, occupation, sex, race, location, and part-time status. Exceptionally well-qualified undergraduate students may enroll in the university's Honors College. The Academic Opportunity Program provides special academic assistance to a finite number of less-qualified but promising students. Pursuit of individual goals by disadvantaged students is encouraged and facilitated. All of these opportunities ultimately reflect and help maintain the standards of excellence of the entire university.

The continuing maturation of Ball State University will lead to a broader definition and assignment of responsibilities for both the faculty and the university itself. While quality teaching will remain the most highly valued focus for the university's faculty, increased attention and resources will be given to research and other scholarly productivity and public service activities. The university will obtain increasing proportions of its financial support from external sources, both public and private. The result will be a balanced and complementary relationship among teaching, research and other scholarly productivity, and public service that will distinguish Ball State from both land-grant research institutions and liberal arts colleges, to which research and public service are less important.

The mission of the university will undergo continuous scrutiny. Significant changes will become necessary as new students, fields, programs, tech-

niques, interests, and technologies affect the institution. Yet the future will not alter the university's firm commitment to the highest possible quality in its activities of teaching, research and other scholarly productivity, and public service, or to the preservation of a positive and attractive environment for learning and living.

Background

When the American Association of State Colleges and Universities announced its intent to assist ten member institutions in evaluating program effectiveness in general education, the offer was seen as a timely opportunity for "fulfilling a deferred commitment." Although the mood among institutions of higher education generally was to move away from the "loose" programs of the preceding decade to some kind of core curriculum, patterns of change reflected little direct relationship to the development of generic skills assumed to be the outcomes of particular programs and, thus, the hallmarks of an educated person. With its special focus on the intellectual tools of analysis, synthesis, communication, quantification, and--possibly--valuing, and with the promise of special assistance in assessing these skills, the AASCU invitation sparked notable interest.

APEP, with its primary emphasis on assessing skills rather than particular subject achievement, carried a special appeal to those on the faculty who felt that all undergraduate instruction should reflect a sensitivity to developing the generic skills of not only communication and quantification but analysis and synthesis as well. Such a view was reinforced by the statement of purpose for general education in the Ball State University catalog, which asserts that the aim of general education program is "to provide undergraduates with the concepts, understandings, skills and values necessary for educated men and women to live purposefully in modern society." Moreover, the statement asserts that "the distinct courses in general education share common goals in the sense that they all should promote the spirit of inquiry, relate knowledge from various fields, and encourage continued liberal education on the part of the student."

With the above institutional philosophy in mind, it was thought that Ball State University's participation in APEP might provide evidence of success in terms of one of the institution's major goals--namely, to provide general education. The nature of the skills to be measured suggested further that the institution might determine whether its program promoted a spirit of inquiry and related knowledge from various fields. The project's concern for assessing such outcomes as the ability to analyze and synthesize was particularly attractive to the many faculty members who recognized the importance of these outcomes but who lacked sophisticated means for their assessment.

Apart from these direct outcomes related to the goals of general education, the university viewed the project as an opportunity to find answers to certain policy questions common to its total undergraduate curriculum. First there were the policy questions of the meaningfulness of separating general education from special-interest education, i.e., of answering the query, To what extent are sharp distinctions warranted between general and specialized programs? It was thought that participation in APEP project would reveal the differences in educational outcomes between those students pursuing highly structured professional progams and those whose major concentrations allowed greater freedom of selection from within the general studies offerings.

A second policy question of timely concern was that of balance between constraint and variety; in other words, to what extent was the program of general studies ensuring the realization of total institutional goals amid increased pressure for greater autonomy for various schools and colleges within the university? At Ball State the general education of students varies widely in relation to use of electives recommended by their respective major departments; therefore, the university was faced with legitimate policy issues concerning centralization and decentralization of curricular and instructional practices. The program of general education, in becoming the focus of these tensions, offered a clear case of policy delineation. One important outcome of the current AASCU-APEP project would be to ascertain the optimal balance of constraint and variety in a program that would also meet standards of effectiveness in the four areas of concern in this evaluation.

Establishing the focus of responsibility for certain educational outcomes was the third policy matter that might be settled through participation in the consortium of AASCU-APEP institutions. To what extent and by what means could the institution place the responsibility for developing communication and quantification skills in all academic units rather than exclusively in the departments of English and mathematics? Institutions with open admissions standards are receiving an increasing proportion of students lacking basic skills of computation and communication. Because current responses to this dilemma have proven inadequate, Ball State University recognizes APEP project as an opportunity to involve a wider segment of the faculty in addressing the problem.

Implementation

Defining Generic Skills

Shortly after it was announced in October 1979 that Ball State University had been chosen to participate in APEP, then Provost Richard W. Burkhardt extended an open invitation to the faculty to attend a general session explaining the intent and goals of APEP. Many faculty members responded to the invitation and were subsequently invited to attend a general organizational meeting.

At this meeting, attended by some forty faculty members, four groups were formed, each charged with developing conceptual and operational definitions for one of four skill areas under study. These four groups continued to meet both individually and in plenary sessions throughout the winter quarter of the 1979-80 academic year, refining and reworking the definitions. Their efforts culminated in a workshop held at the University Kitselman Center in February 1980. During the workshop, much discussion focused on defining the levels of performance for each of the four abilities and on planning an approach to identifying when and how these abilities were being taught in the general education curriculum. Throughout March and April of that year, faculty members continued to meet in general sessions, improving not only on the conceptual but also the operational definitions. Six basic assumptions informed and guided the thinking of the four committees:

- that the four thinking skills defined were but a few of a wider array of important skills and that definitions can be developed only in such a context (for example, analysis and synthesis are only two of six for the Bloom taxonomy)
- that thinking skills can be defined in behavioral terms and observed, described, and measured in some way
- that given the present constraints of the state of the art of educational measurement, the definitions, in order to be useful in this project, need to be susceptible to "pencil and paper" measurement
- that the context adopted by the Ball State University faculty for defining these skills is primarily cognitive rather than affective
- that such skills can be defined more distinctly at basic levels and that definitions become interrelated at advanced levels
- that such skills present infinite possibilities for development throughout life.

By the end of April 1980, each of the four committees had settled on the following basic conceptual definitions.

Communication

The committee on communication defined this ability as receiving and sending messages through such processes as reading, listening, writing, and speaking. The committee pointed out that although nonverbal communication was not included, its importance was not to be denied. Furthermore, it was noted that the component skill of effective speaking was perhaps the skill least susceptible to the type of assessment associated with APEP.

Analysis

The committee on analysis developed the following statement: analysis is the process of separating a communication into its constituent parts to make clear the order of and relation among ideas expressed; to indicate the organization of the communication and the way communication manages to convey its meaning; to show the bases for and arrangement of the communication, both explicit and implicit. One issue that surfaced in constructing the definition of analysis involved the notion that there is a context, not developed in the definition presented above, in which the skills of analysis and synthesis are complementary components of the broader skill of "problem solving." A question that also promoted significant discussion was that of whether it was possible to widen the scope of the skill of analysis beyond a "communication" and still render the skill susceptible to measurement in the traditional sense?

Synthesis

The committee dealing with synthesis presented its conceptual definition by identifying six operational components:

- to organize unrelated facts and ideas into a framework
- to introduce new ideas into the perspective of an existing framework
- to speculate on new frameworks within existing systems
- to extrapolate from known to unknown situations
- to recognize connections between theory and practice
- to integrate new facts and ideas with personal experience.

Two unresolved questions emerged from the deliberations concerning the skill of synthesis. One was whether the parameters of APEP reduced the notion of synthesis to verbal communications. A second was that of whether the skill of synthesis should be defined in a way that complemented the definition of analysis.

Quantification

The quantification committee, after much discussion, accepted the following definition: the ability to understand and use symbols and diagrams to express quantity and relationships. In its discussions on quantification, the committee wondered whether the definition finally determined was too closely associated with the discipline of mathematics. Additional musing led to the suspicion that advanced quantitative skills could well be very similar to the skills of analysis and synthesis.

Assessment of Skills

Upon acceptance of the operational definition of skills by the participating faculty members, a committee was selected from the group and charged with the task of constructing a research design for the project. A preliminary evaluation design emerged in spring 1980. A pilot program was proposed to test the validity and applicability of test instruments sent to the university by the Educational Testing Service (ETS) under the direction of Jonathan Warren. With a final evaluation design for the project approved by the committee in early October 1980, the assessment of skills was undertaken.

Purposes

As the sequential stages of the APEP paradigm unfolded, the general purpose of the project was increasingly refined. The unique nature of the general education program at Ball State University together with the time constraints of the APEP project caused the Ball State University planning group to view the initial implementation year of the project as a formative step in continuing evaluation of the general education program. The purpose of the first implementation year was articulated as follows:

> To assess entry-level and exit-level student performance in four generic skills (i.e., analytic, synthesizing, quantitative, and communication) to initiate a program of continuing evaluation of the effectiveness of the Ball State University general Education Program, as it functions in concert with majors in the major curricular divisions of the University (i.e., humanities, social and behavioral sciences, and sciences and mathematics, or business, technology, and applied fields).

To achieve this overarching purpose, Ball State made the following decisions on the nature, scope, and focus of the assessment mission:

- to focus the investigation on the four generic skills mentioned above and to omit the generic skill of valuing
- to use criterion-referenced (competency) measures to assess high-, middle-, and low-levels of skill performance
- to adopt and validate available instruments that were relatively "discipline-free"
- to supplement the basic measures with various unobtrusive data such as entrance test results and achievement of various levels of academic honors
- to focus the first-year implementation study on outcome assessment of both freshmen and seniors
- to use a limited program portrayal approach for the assessment of skill growth of freshmen who would take various combinations of five or six general education courses during their freshman year and to study the difference between seniors who pursued a structured program of general education and those who did not do so.

It was anticipated that this approach would facilitate a number of significant outcomes: (1) a provision that would allow judgments about the entry and exit generic skills of current Ball State University students as well as a basis for making some limited judgments about the effectiveness of certain combinations of general education courses, (2) judgments about the validity and scope of the various measures used, and (3) a data base not currently available to facilitate future evaluation efforts. It was also hoped that this approach would stimulate interest in carrying out increasingly focused outcome and program portrayal evaluations in subsequent years.

Types of Skill Assessment

Standardized, multiple-choice instruments prepared, field tested, and revised by ETS were chosen to measure the four generic skills. The specific forms for communication, analysis, synthesis, and quantification were ETS Form 14 Revised, ETS Form 34 Revised, ETS Form 44 Revised, and ETS Form 27 Revised, respectively. These particular forms were chosen from three of four alternate forms for each skill by the research design committee, which hoped to reduce the total testing time to one hour. The committee did recognize that the brevity of these instruments might raise questions about their power, reliability, and validity.

Specific Cohorts Selected for Testing

Five cohorts of students were defined for testing in the assessment stage of the project: (1) 375 matriculating freshmen, (2) 91 of the 375 freshmen seven months after matriculation, (3) 260 randomly selected seniors in their final quarter of studies, (4) 39

seniors who had been admitted in the "distinction" category and who had completed the structured Honors College curriculum of general education, and (5) 44 seniors who likewise had been admitted in the "distinction" category but who had not pursued the honors curriculum. With the exception of the matriculating freshmen, who were tested in the autumn quarter, all other groups were tested late in the spring quarter of 1980-81.

Because the major thrust of the project was descriptive and analytical, the following questions were viewed as generally more helpful to the evaluation than specifically formulated hypotheses:

- Are there distinct differences in skill performance in freshmen after a year of general education?
- Are there differences after a year between those freshmen who complete a substantial number of general education courses and those who do not?
- Are there distinct differences between matriculaters and graduating seniors?
- What proportion of the beginning freshmen have attained *low* skill performance, *medium* skill performance, and *high* skill performance in each of four areas?
- What proportions of the seniors have attained *low* skill performance, *medium* skill performance, and *high* skill performance?
- Are there significant differences in seniors of equal ability between those who pursue a structured general education and those who do not?

Study Results

Communication

On the test of communication skills the freshmen showed some gains between the autumn and spring quarter, but the differences were not significant at the 0.05 level of confidence. A formal analysis of scores made by freshmen who had completed eight or more general studies courses and those who had completed five or fewer such courses showed a slight difference between groups, but not significant at the 0.05 level of confidence. When beginning freshmen were compared with seniors, the latter group reflected a higher skill level. The differences were found to be significant at the 0.001 level of confidence. When arbitrary standards were set for low, medium, and high performance for the communication test, the range and distribution of scores for freshmen and seniors reflected the differences reported above. A comparison of differences between two groups of seniors--those completing the more structured general education program of the honors curriculum and an equally able group not pursuing that curriculum--showed the former group had higher scores than the latter. However, differences were not found to be significant at 0.05 level of confidence.

Analysis

On the test designed to measure the intellectual skill of analysis, the freshmen showed statistically significant gains between the autumn and spring testing. The limited sample (N=28) of freshmen having taken eight or more general studies courses was compared with those taking only five or fewer (N=29). The group having completed five or fewer general studies courses scored significantly lower than the group completing eight or more at the 0.001 level of confidence. On this test seniors performed better than freshmen, the difference being significant at the 0.001 level of confidence. Again, when arbitrary standards were set for various levels of performance, the gains of freshmen and the differences between freshmen and seniors were obvious and consistent. When the two groups of academically able seniors were compared, the average mean score of the Honors College group was markedly higher than that of their classmates of equal ability. Differences were found to be significant at the 0.001 level of confidence.

Synthesis

On the test designed to measure the skill of synthesis, gains made by freshmen between the autumn and spring were significant at the 0.001 level of confidence. The informal analysis of two groups of freshmen with varying experiences in general education showed the group with more such courses to have only slightly higher increments than its counterpart with fewer general courses. Seniors, however, showed significantly higher scores than did freshmen on this test. When the range and distribution of scores were arbitrarily grouped in reference to low, medium, and high standards of performance, the differences both between freshmen groups and between freshmen and seniors reflected the earlier findings. Again the Honors College seniors performed better than their classmates of equal ability. However, differences between groups were not found to be significant at the 0.05 level of confidence.

Quantification

Of the four skills measured in the study, quantification reflected the greatest difference within the freshmen cohort and between the freshmen and seniors. Differences were significant at the 0.001 level of confidence, and t-values were the highest of the four

areas of comparison. Statistical comparisons showed the group of freshmen with five or fewer general studies courses to have a lower average mean score than that of the group with eight or more such courses. However, differences were not significant at the 0.05 level of confidence. Differences between average mean scores of the Honors College seniors and their counterparts were noticeable but not significant at the 0.05 level of confidence.

Conclusions

To find significant incremental differences between the freshmen tested in the autumn and those retested in the spring on three of the four tests is cause for general satisfaction with the experience provided students in the initial year of studies. Without national norms for the tests administered, though, it is difficult to know whether the increments are sufficient to make positive evaluations of that experience. Furthermore, the limited samples of students having varied contact with general courses precludes anything beyond conjecture that such courses are more likely to develop the skills of concern in this study. The finding that the differences of test scores on the communication test were not significant provides little help in assessing the value of freshman English composition courses. The rather limited nature of the test, which emphasizes reading comprehension to the neglect of other important aspects of communication, points out rather specifically the need for identifying and using a more adequate test for measuring this important skill.

In contrast, the test for quantifying skills was perceived to measure the skills specifically sought in the introductory courses in mathematics. The data related to performance on the quantification test, on which differences were perceived to be most significant, also present a quandary. A formal analysis of the groups with varying numbers of general courses suggests that the panel of general studies courses have little more impact than would a collection of unspecified electives.

As with the increments shown by freshmen, the favorable differences achieved by seniors over the freshmen suggest the positive impact of the undergraduate experience. Again, comparison with normative data would have provided greater assurance that the differences were sufficient to warrant a high level of satisfaction with the outcomes. Also, it is not known how much the difference might be attributed to normal maturation. The fact that on all four tests the increment shown in the freshman year is equal to the differences between freshmen in the spring and seniors may suggest a leveling off of the development of these skills as the student progresses through four years of undergraduate studies. This diminishing rate of change should at least be examined for its fuller meaning.

Probably one of the more provocative questions to be studied in interpreting the data from test scores deals with the differences found between the two cohorts of seniors who had been admitted four years earlier in "Distinction" category. A control group of forty-four seniors in this category had pursued a general studies program characterized as having little structure and consisting of electives in the various distribution areas of humanities, social sciences, and natural sciences. In contrast, the group of thirty-nine seniors who were completing the Honors College curriculum had pursued a more structured program of general education. This program comprises two values-oriented symposia in the first year, a three-course humanities sequence in the second year, two informal colloquia in the third year, and a capstone honors thesis in the last year. On all those tested, the honors group performed better than their classmates with comparable entrance test scores. Further analysis of these differences could provide the investigators with significant information on which to base assessments of this particular pattern of general studies.

Policy Development and Decision Making

Three fundamental policy questions confronted the university at the outset of this study: (1) To what extent are sharp distinctions warranted in separating general education from specialized studies? (2) How should a program of general studies be structured to achieve a balance between constraint and variety? and (3) Where should the university place the locus of responsibility for developing and appraising the program of general studies? Participation in APEP provided increased insight into these pervasive issues and generated a greater level of faculty consciousness of the need to make a series of recommendations in response to their new understanding of the comprehensive problem of improving general education. This section of the case study summarizes the nature and scope of this heightened consciousness and proposes those general and specific policy recommendations which seem in order.

Response to Basic Issues

The question of maintaining a distinct separation of general and specialized education forces educators to reconsider the overall purposes of undergraduate education, the unique claims made for common

learnings, and the appropriate emphasis on vocational preparation. There was nothing in the experience of the Ball State University faculty task group to suggest that the aims of general education would be served automatically in an undergraduate curriculum increasingly characterized by specialized education. Instead, the discussions on the campus in small groups and plenary sessions supported not only the concept of a visible program of general studies but also one that would foster the development of those intellectual skills that were the particular focus of this study. In addition, there seemed to be full recognition of the need for making general education a university-wide enterprise of concern to all, even though primary responsibility for its delivery may rest with the faculty of particular disciplines. What was needed, it seemed, were better mechanisms for ensuring that what was provided in the name of general education was indeed all that it promised to be. Many of the recommendations offered by the group emanated from the conviction that the cause of general education should not be abandoned but instead reinforced consciously by the faculty and administration. This would require greater attention to the structure of the program, to quality review by various governance units, and to continuous evaluation by an administrative unit with resources sufficient for the task.

The issue of structure in an ideal general studies program pervaded the deliberations of the faculty throughout the period of this study. Whatever consensus achieved was limited to a general feeling that the existing program lacked sufficient structure to achieve its purposes. A majority of participants considered the surfeit of existing options so amorphous that a single program for all students was difficult to identify. When questions were raised about the parts of the curriculum that develop generic skills, it was apparent that, with the exception of the skills of communication taught in two required English courses, there was no component of common learning; on the contrary, each student's program of general education constituted a mixture of conveniently scheduled electives, directed electives in support of the majors, and whatever additional courses might be needed to fulfill distribution requirements. This limitation was particularly obvious at that point when the group wanted to compare the performance of several student cohorts with different patterns for meeting these requirements. There were as many patterns as there were students and, with the exception of the Honors College seniors, no consistent patterns prevailed.

Although no formal proposals on structure emerged from the APEP experience, there seemed to be support for a general education program characterized by a common core of specific courses complemented by electives from specified groups of courses in the broad fields. Evidence from the study of skills developed by seniors who had participated in a structured sequence of Honors College courses supported this inclination. As in most discussions of curriculum structure, there was an articulate minority who eschewed any required formal organization but instead brushed aside the issue of structure for an emphasis on the instructional strategies within courses.

Inasmuch as the university launched this study with some concern about the locus of responsibility for developing and evaluating the general studies program, it is appropriate at this point to assess the experience in this regard. Apart from the general argument that such studies, and particularly the development of generic skills, should be a responsibility shared widely in the university community, there was wide recognition of the need for regular and continuous assessment of program and student performance. A review of some of the major assessment problems will illuminate this need in greater detail.

Shortly after the APEP task force began to work through the various stages of the evaluation paradigm, it became apparent that the university did not possess in a single administrative unit all the resources needed for pursuing a project of this scope. Although responsibility for conducting the project rested nominally in the Office of Undergraduate Programs, the sources of expertise in measurement, design, and statistical analysis were located elsewhere within the university's organizational structure. Furthermore, an emphasis on voluntary participation, made primarily because the institution was undergoing rather frequent turnover in top administrative leadership, meant that the task force included for the most part faculty members and administrators who, though committed to the project, lacked certain necessary skills in evaluation. The experience of dealing with these deficiencies has led to the conclusion that the university needs to place clear responsibility for this type of assessment in a single administrative unit possessing the necessary staff and inclination to conduct such studies on a routine and continuing basis.

To establish an assessment center whose responsibilities would be to assess and certify student attainment of the generic skills, the university would need to reorganize existing units that share but do not claim full authority or responsibility for these matters. Presently the university maintains the

following units, which collectively possess the necessary staff: Office of Institutional Studies, University Computer Center, research computer design unit of the Office of Research, Director of Analytical Studies, the psychometric staff of the Office of Counseling and Psychological Services, and the Director of Student Affairs Research. These offices, however, are administered by the president, provost, and the vice president for student affairs. Should the university decide against reorganizing its resources into a new unit, there would need to be an unmistakable mandate from the chief executive officer clearly identifying the responsible coordinating unit and the cooperation necessary from other support units.

The choice of instruments for use in this study represented a compromise between aspiration and expediency. Although the limitations of the scales developed by the ETS were readily recognized, the will to develop better ones within the available time frame was lacking. After examining the COMP materials produced by the American College Testing Program and those of the McBer Company, the faculty opted for the shorter scales under development by ETS. The choice of these instruments was predicated on the anticipation that national norms might be made available during the course of the project. Unless an institution has sufficient facilities and faculty interest to develop "home grown" measures that fully reflect their objectives related to generic skills, there is little choice except to use currently available commercial instruments. Many in the group saw in this dilemma an additional area of professional development that would be served by the university's Office of Instructional and Professional Development.

Although certain compromises were made in selecting the various samples of students to be tested, subsequent analysis suggests that fairly representative cohorts were included. The Research Design Committee would have preferred to test the entrance skills of matriculates during the orientation process in summer 1980 but had not progressed sufficiently toward a selection of measurement instruments to accomplish the task by that time. Consequently, it was necessary to choose as test sites those sections of English and mathematics courses with high enrollments. Similarly, the choice of courses with a high enrollment of seniors as sites for exit testing was seen as less desirable than a pattern of testing that would identify clearly a representative sample of students to whom a stipend would be paid for their cooperation. In contrast, the voluntary participation without stipend by third-quarter seniors in the Honors College seemed to reflect the most optimal conditions for assessing performance on the several tests. An institution that proposes to measure generic skills at the close of the undergraduate experience would be advised to stipulate such participation by students in graduation requirements; otherwise, the sampling aspect will continue to present problems.

One of the unexpected outcomes of the APEP experience at Ball State University was the increased sensitivity to the need for determining in what areas students might acquire the skills under study. The faculty expressed a particular concern for identifying the intentional steps instructors take to reinforce skill development. This expression was seen most obviously in the work of the General Studies Committee. During the period of APEP evaluation the committee, many of whose members were active in the project, began an intensive review of the program of general education. Its dialogue with departmental representatives, examination of course syllabi, and other program review activities led the committee and APEP participants to conclude that insufficient attention was being given course objectives and instructional practices bearing on the development of intellectual skills. Apart from the information provided on student and program performance in relation to the four generic skills under study, this recognition of the need for curriculum improvement may have been the single most important outcome of the study.

Recommendations

On the basis of the experience given a relatively large sample of the faculty and staff in the past two years, it seems appropriate to recommend the following:

- that the Undergraduate Educational Policies Council develop a policy statement that commits the university to a continuing program of assessment of student performance in the skills of communication, quantification, analysis, and synthesis, with special attention to entrance and exit level of skill performance
- that the university officers determine which administrative unit should serve as an assessment center and provide the locus of responsibility and authority for a continuing program of student assessment, and that they provide for such a center the cooperative use of necessary resources such as specialists in research design, psychometrics, computerization, and data analysis
- that the General Studies Committee in its current planning for a revised program of general education give special attention to the need for greater

structure in the program; course approval criteria that reflect the importance of course objectives and instructional strategies related to generic skills; and continuous evaluation of the program

- that both the Department of Educational Psychology and the Department of Secondary, Higher, and Foundations of Education encourage as a component of their graduate studies increased research of problems, issues, and practices related to undergraduate instruction and curriculum development at this institution
- that the Office of Instructional and Professional Development provide, through workshops and other means, specific assistance in developing course objectives and instructional strategies related to generic skills.

Written By

Thomas A. Kaluzynski and Victor B. Lawhead

Institutional Contacts

Thomas A. Kaluzynski
Victor B. Lawhead

California State University, Chico

Institutional Profile

California State University, Chico began in 1887 as Northern Branch State Normal School of California and is the second-oldest school in the current California State University system. Its purpose was "the training and education of teachers in the art of instructing and governing public schools of the state." As might be expected, Chico State has changed dramatically from its founding status of five faculty members and ninety students. Other colleges and universities began to accept appropriate work in liberal arts courses at Chico for transfer credit in 1913. Chico State Teachers College was first authorized to confer the Bachelor of Arts degree in 1924. Master's degrees were first offered in 1949, and in 1971 the institution was formally titled "university."

At the present time, Chico State can most accurately be described as a large liberal arts university whose primary mission is instruction. A wide variety of other activities, ranging from basic research to community service, also characterize the university's operations today.

Although it is risky to predict the future, it seems fairly certain that the university's future mission will be less broad than that pursued in its past. Without major responsibility for either two-year programs or research functions, it is anticipated that Chico State will grow in a variety of directions addressing the needs of California's population and developing in a manner commensurate with new understandings of educational structure and process. No future development is expected to change either the basic instructional mission of the university or the faculty commitment to instructional excellence.

Brief statements cannot fully describe California State University, Chico, but the following characterizations will certainly provide the reader with a good overview:

- a geographic service area of twelve northeastern California counties, which is about one-fifth of the total area of California, and larger than ten of the fifty states
- a remarkably beautiful campus of more than 100 acres, augmented by an 800-acre university farm and numerous agreements providing for the conduct of university instruction at a variety of off-campus sites throughout its service area
- a faculty of about 700, the majority holding doctoral degrees, with a similar-sized support staff; of the latter group, about 45 percent working in direct support of instruction (instructional, administration, library, and so forth)
- a curriculum covering most of the major academic subdivisions of knowledge. Only architecture, law, library science, and theology are absent. Baccalaureate degrees are awarded in more than fifty areas of study, and about half of these offer master's degree programs as well
- a student body of about 14,000. Of these, more than 80 percent attend full time. More than 85 percent are undergraduates. The sexes are equally represented. About 40 percent of the university's enrollment is new to Chico State each year, and of these, about half are transfer students from other (mostly two-year) colleges and universities.
- few students are academically disqualified--perhaps two or three percent each year--and relatively few students migrate to other colleges and universities. About 70 percent of our freshmen return to continue their education in the sophomore year.

In 1973, the former Federation of Regional Accrediting Commissions of Higher Education received a grant from the Danforth Foundation for a three-year study, "Looking Toward the Development of New Techniques for the Evaluation of Institutions of Postsecondary Education Which Would Attempt to Emphasize Outcomes Measurements." Chico State was a participant in that study, and a description of its involvement and accomplishments is reported (under the pseudonym of Big Creek University) in *Evaluation of Institutions of Postsecondary Education: Assessment in Terms of Outcome Through Institutional Self-Study.* Chico State gained much valuable experience from that study, of use in improving both its educational planning and ability to conduct self-appraisals.

Background

In addition to the institution's involvement in the self-study experimentation described earlier, Chico State has a unique history of evolution in its general studies program. Beginning in 1972, the faculty, in partnership with the academic administration, ad-

dressed the issue of establishing general studies goals and building a program based on those goals. After establishing criteria that supplemented the goals, an ad hoc committee created by the faculty senate spent three years reviewing every course then listed as satisfying general education requirements. They met with faculty members and school administrators and formulated a new list of recommended courses. In the end, the general education list was pared from some 2,000 to 250 courses. Although there was understandably some unhappiness with individual decisions, there was also a general acceptance by the faculty and administration that the job had been well done and that the university had the beginnings of a sound general education program. The ad hoc committee was replaced by a faculty advisory committee, and the undergraduate dean in the office of the vice president for academic affairs assumed the role of general education coordinator.

The general education program has consistently been addressed in the university's planning process, and it is generally recognized that the program must be protected from course proliferation and narrow departmental resource concerns. The faculty senate has been highly protective of the program and consistently resists efforts to turn general education into a political issue.

Chico State became involved in the APEP program at a point in its history that was both fortuitous and problematic. The university had, up-to that point, seven years of experience in developing a general education program in terms of a set of goals generated and approved by the faculty. Many faculty members however, felt that some revisions were in order and a new phase of program building was appropriate. Furthermore, a modest process evaluation of the existing program had led to additional questions about the impact of general education, as measured by differences in the levels of accomplishment on achievement tests of native and transfer students as well as of freshmen and seniors.

During this period a California State University system task force was working to develop a new proposal for the California Administrative Code segments on general education. University officials were aware that the trustees would act on this matter by May 1980 and subsequent to that action the campuses would have approximately one year to make their programs fit the new guidelines. Drafts of the system's task force materials were circulated for campus response, and it became clear that the new guidelines would be far more prescriptive and detailed than ever before. The proposal for participation in the APEP project was written in this context and placed major emphasis on gathering data from students who would complete the existing general education program. In this way, data would presumably be available for drawing comparisons between their accomplishments and those of students completing the program to be initiated in the fall of 1981. In addition, freshmen/senior differences and the differences between native and transfer students would be measured.

Implementation

Research Design

The research design for the study is given below in the form of four basic questions asked about any combination of goals for which measurements were obtained. The narrative makes references to transfer and native students. "Native students" are defined as those freshmen entering Chico State with no transfer work from another college or university. "Transfer students" are those with fifty-six or more units from other colleges or universities accepted at Chico State. Students with intermediate amounts of transfer work were not incorporated in the design.

What is the level of accomplishment of Chico State students?

Two groups of thirty randomly selected native seniors were tested in the fall semesters of 1980 and 1983. Some scores of these groups compared with national, state, or local norms for some of the tests being used.

Is experience with general studies and other courses at Chico State correlated with changes in the level of accomplishment?

An additional group of 100 randomly selected students who had been tested as freshmen in the fall of 1980 were tested again in the fall of 1983.

The above question was answered by comparing the scores of the 100 seniors with their same scores as freshmen. Freshmen-versus-senior differences were assessed both within 1980 and within 1983 to detect how students change during their college years.

When do changes in the level of accomplishment take place during a student's career at Chico State?

A separate, randomly selected group of thirty

native freshmen were tested each fall semester from 1980 through 1983.

A comparison of these four groups of scores was intended to provide information relevant to detecting the point in time at which changes occur and the effects of repeated testing. Unfortunately, student participation problems reduced the sample sizes, and attention to this question was deleted from the study.

Is there a difference in the level of accomplishment of native and transfer students?

Two additional groups of thirty randomly selected transfer seniors were tested--one in the fall of 1980 and the other in the fall of 1983. The scores of these transfer senior students were compared with the scores of native seniors.

The Role of APEP in General Education Program Development

Faculty on the campus were aware that by May 1980 there would be new system guidelines for general education, and that there would be about one year to develop local implementation policy. Clearly, neither the Educational Policies Committee of the Academic Senate nor the General Studies Advisory Committee was willing to spend any time on any revision or rethinking of the general studies program until the system trustees had acted. We of the project staff knew that once this radical redistribution of general education content was announced, campus interest and participation in formulating local policy would be intense and pervasive. At this time, we made several proposals to the president and to the vice president for academic affairs. All the proposals, listed below, were subsequently accepted.

- In the fall of 1980 we would construct an interim program for the 1981-82 academic year. This interim program would provide time to develop a more thorough approach to revising our policy;
- The 1980-81 academic year would be devoted to a thorough and careful restudy of campus general education goals in the light of system policy and the development of local policy for program building, monitoring, and evaluation, as well as criteria for course review;
- The APEP generic skills would be systematically embedded into the new criteria;
- All courses proposed for the new program would be carefully and individually reviewed by a faculty committee using course descriptions, sample examinations, and sample assignments, as well as criteria for course review;
- By the end of the fall semester 1981 all but the upper-division program would be in place and by April 1982 the whole program would be complete;
- Every five years one-fifth of the programs would be rebuilt *de novo*, with each course reviewed in terms of the general studies criteria. We realized that this timetable was out of harmony with the sequence of the APEP paradigm as it evolved.

At that point, we had to work with the existing general studies goal statement as our conceptual definition. It was decided to use a two-committee structure. The governing committee for the APEP project was the long-standing advisory committee on general studies, whose charge was to advise the vice president for academic affairs, or his designee, the undergraduate dean. This committee was to have the technical and advisory support of a committee whose members were faculty members with expertise in tests and evaluation and who came from a variety of academic areas. The university's test officer and the director of institutional research served as staff to that committee.

- Student accomplishments in general studies be regularly measured and the outcome information provided to the appropriate committees. Using the criteria stated in our general studies goals and those of APEP, the evaluation committee prepared a series of proposals for review by the General Studies Advisory Committee. After several consultations and subsequent revisions, a final list of evaluative instruments was selected.

Test Measures

All student participants took a thirty-minute survey titled the *Student Information Form* (SIF; Cooperative Institutional Research Program). This survey provided demographic and attitudinal information valuable for later analyses of the outcome measures. All students also had a university record of their grade-point average, workload, etc., for each semester under study. The remaining measurements are described in sections identified by their focus on local or APEP goals or both:

I. Literacy (CSU, Chico's Goals of General Studies); Communication (APEP); and Quantification (APEP)

A. *A Student Composition* writing task constructed locally. This was a sample written under "compare and contrast" instructions such as those used in the state's English Placement Test and English Equivalency Examination. Local holistic scoring was

conducted by experienced readers from the English Department faculty, and the results were compared with state norms for students in the CSU system.

B. The *Cooperative English Test* published by ETS. Two sub-scores titled Reading Comprehension and English Expression were obtained from this standardized, objective test having national norms.

C. *Measures for Academic Competencies in General Education* available from Jonathan Warren of ETS. There were two of these tests, one in communication and the other in mathematics.

D. *National Assessment of Educational Progress (NAEP): Mathematics* published by the U.S. Office of Education. This test required students to use numbers, presented in digital, figurative, and narrative form, in problem-solving tasks. National norms were available for 17-year olds and for young adults (18 to 35) but not for college students.

II. Critical Reasoning (CSU, Chico Goals of General Studies); Problem-solving (CSU, Chico Goals of GS); Analysis (APEP); and Synthesis (APEP)

A. *Test of Thematic Analysis* published by McBer. It asked students to identify and describe the differences between two groups of stories. High scores represented a differentiation and evaluation of story content; low scores represented differentiation based more on the reader's personal response.

B. *Analysis of Argument Test* published by McBer. Students were asked to produce arguments on both sides of an issue. Scores represented the ability to organize and evaluate arguments about both sides of an issue.

C. Rokeach *Dogmatism Scale.* This test was not published but exists in the research literature of social science. The test was designed to assess the openness of an individual's belief system.

D. *Measures for Academic Competencies in General Education* published by Jonathan Warren of ETS. There were two of these tests, one in analysis and the other in synthesis.

III. Artistic Expression (CSU, Chico Goals of GS)

The College Level Examination Program (CLEP) General Examination iin Humanities) published by ETS. Subtest scores in Fine Arts and Literature were obtained and scored in reference to national norms.

IV. Form of the world in which we live (CSU, Chico Goals of GS)

The *CLEP General Examination in Natural Science* published by ETS. Subtest scores in physical science and biological science were obtained and scored in reference to national norms.

V. History and Government (CSU, Chico Goals of GS)

The *CLEP General Examination in Social Science-History* published by ETS. Subtest scores in social science and history were obtained and scored in reference to national norms.

VI. Aspirations and problems of subcultures (CSU, Chico Goals of GS)

No tests were proposed for this goal.

VII. Influence of technology (CSU, Chico Goals of GS)

No tests were proposed for this goal.

VIII. Integration of major work (CSU, Chico Goals of GS)

No tests were proposed for this goal.

IX. Valuation (APEP)

AVL Study of Values published by Houghton Mifflin. Six basic interests were assessed with this instrument: theoretical, economic, aesthetic, social, political, and religious. National norms were available.

The above list was the result of substantial effort. Faculty committees reviewed a great amount of pertinent literature, met regularly, and debated and discussed the issues with care. They found it to be an extraordinarily complex task and one wherein no single result could be defined without some ambiguity. Despite structural difficulties, the proposed design was submitted with the confidence that it represented substantial progress toward several ends, namely:

- This proposal defines several ways in which better information on Chico State's accomplishment in general education can be acquired;
- It was consonant with APEP's orignial definition of project tasks;
- It was consonant with the history and current conduct of self-study related to general education at Chico State and at the CSU system;
- The information acquired from this study would stimulate greater and continuing faculty involvement and dialogue in matters related to general education.

The testing was completed on schedule. The participation rates of students were lower than we had hoped even though sample students were individually contacted and paid for their time. However, the first phase of the testing program was completed with adequate numbers and the baseline data needed were generated. That done, we returned to stage one as, in the late fall 1980, the faculty was now ready to address the general studies criteria in response to the system-mandated guidelines.

Campus Response to New System Guidelines for General Studies

As mentioned above, the Educational Policies Committee of the faculty senate was charged with

developing a policy recommendation for the entire academic senate to consider. This strategy included receiving proposals from the various academic units on the campus with ideas for criteria. A subcommittee of the Educational Policies Committee developed early drafts of some of the broad conceptual statements. The Educational Policies Committee as a whole reviewed these conceptual statements as well as the more specific criteria. The project was given the broadest possible publicity, with people from the various academic areas invited to attend the meetings as proposals pertinent to them were discussed.

Our intention was that the process be open and the widest possible participation be solicited because we knew that the redistribution of general studies was going to have a significant impact on all units of the campus. Open hearings were held and subcommittees wrote policy to incorporate conceptual statements and explicit criteria in their work. Many revisions, discussions, compromises, and new documents emerged. A draft policy document was sent the faculty senate and was also distributed widely in the campus community. Sessions of the faculty senate were well attended while the new general education policy was debated, and in May 1981, with relatively few revisions, it was accepted by the president.

The skills for APEP, as well as other local goals, were written into the policy. Fall of 1981 was devoted to the selection of courses for the program by faculty review committees established under the policy. The total number of courses was limited to 200, so the issue was not merely which courses met the criteria but, rather, which courses best met them. The burden was on faculty members to demonstrate that not only did their courses cover the content area, but also communication, particularly writing skills (in all courses), quantification (where appropriate), critical thinking, analysis, and synthesis. In the upper-division program, course proposals had also to demonstrate an integrative component and a focus on values. Examinations and assignments for all courses had to be consistent with course outlines.

The lower-division program is now complete and the upper-division program was completed in April 1982. Our next task is to review the testing program to determine whether its content is still consonant with the program criteria. The evaluation committee will be reconvened for that task. The next large testing by Institutional Research with standardized instruments will have taken place by 1983-84. The previously tested freshmen, now seniors, will have completed the old general studies program; other freshmen will have begun under the new one. Ultimately, a useful evaluation of student achievement changes will result. These changes can be assessed for both pre- and post-APEP times and will likely be the stimulus for a major review or progress report on general studies in the mid-1980s.

The Present Impact of APEP

The impact of APEP at Chico must be viewed in relation to other attributes of our history and plans. APEP merely augmented what was on going and planned. APEP, at Chico, was definitely not a separate endeavor with well-defined initiation and termination dates. Chico has been paying special attention to APEP-like topics for more than a decade and intends to continue that attention. APEP, per se, provided welcome help, additional rationale, and other benefits to this work. A brief description follows.

One result to date is the acquisition of information related to the level of measured achievements by Chico State students. In general, Chico's results were simultaneously informative, reassuring, and insufficient. We have been able to compare some of our test results to those published as norms in test manuals or to study results such as those reported in *The Impact of College on Students* by Feldman and Newcomb. We have not felt confident, however, that we had access to information representative of college student achievements in AASCU institutions or any other well-defined population. The problem is mostly that such information does not exist.

Representative information could not exist at Chico, where almost half of a well-defined sample of students declined to participate. Also, information representative of national college student populations is usually absent and often characterized by sampling flaws when present.

A second result to date is the acquisition of information related to the mean achievements of Chico student subgroups. We expected to find differences between our fall 1980 freshmen and seniors. We were curious about the differences between native and transfer seniors in fall 1980 and in spring 1984. In general, the results to date have been informative and somewhat reassuring. No other APEP institution reported finding significant group differences in the hypothesized direction on all the tests used, and Chico State also could not report such a result. Test information such as that described here is less vulnerable to the sampling criticisms previously mentioned. The absence of defensible norms is almost irrelevant, and the bias resulting from incomplete

local participation is somewhat amenable to additional analysis.

A third result to date involved some of the APEP tests on written communication and mathematics. The communication test results were compared with those of a similar study in 1975 and helped convince the faculty committee on writing that more stringent measures were needed. The committee recommended, and the president agreed, that all junior local students, native and transfer, had to take a screening test in written English. Those not passing at an acceptable level would not be permitted to continue their upper-division writing work until the deficits had been remedied. This policy began in the fall of 1982. Similar attention is being devoted to mathematics. A fourth result to date was the acquisition of program portrayal information of two kinds. A survey of general-education class students and their instructors produced a description of the amount of attention given to each general education goal in classes satisfying general education requirements. We found some classes neglecting some goals and some discrepancies between instructor descriptions of what was delivered and student descriptions of what was received. This informative feedback generated some valuable dialogue.

The second set of information produced has yet to be published. This is a description of the amount of different kinds of course work taken by students in different majors. Those familiar with the literature of higher education management systems may recognize its name: "Induced Course Load Matrix" (ICLM). This information describes the impact of a university's new general education policies on the curriculum consumption of its student body. It also reports student surrogate outcomes from the perspective of quantity of schooling. These course consumption or ICLM reports will be regularly published starting with the fall semester 1983. Historical reports will also be available at the same time.

Some of the overall results described above depended on completion of the testing facet initiated in fall 1980. Other aspects of Chico's program use that information but are not dependent on it. One such program is the University Planning Data Base (UPDB). There we have been working with a computer data-base management system called Scientific Information Retrieval (SIR Inc., Skokie, Illinois), which has great promise for managing the work of a university intending to develop an analytical management information system. The UPDB should be fairly functional by the end of the year. It will describe resource utilization and outcomes, by program, in a manner that is highly informative and exceptionally easy to use. Chico's involvement in the APEP project has helped bring student outcome information into the UPDB, and this we regard as particularly meritorious.

Conclusions

APEP has proceeded, albeit at a snail's pace, between cessation of active supervision by AASCU and the present time. The UPDB is regularly publishing educational statistics for the university community, and post-tests have been administered to those original APEP students willing to participate. Student outcome measurements are now obtained annually. Approximately one-fifth of each year's seniors are sampled and offered both R. Pace's survey measuring the "Quality of Student Effort (Higher Education Research Institute, UCLA) and a locally developed survey of "Educational Satisfaction." All degree recipients are offered a "Survey of Recent Graduates" about four months after their graduation. Many post-APEP research reports are still possible and still intended. They will be published as time for them is available to the institutional research office.

The advent of APEP and Chico's involvement in the project continues to promise a unique benefit to the university. The work and faculty involvement associated with APEP succeeded in "defining" a stand-alone operation that provides informative feedback on the university's most important goal, student achievement(s). Its cost was modest--about 1/1000 of the institution's annual budget--and its nature sufficiently flexible to accommodate all except those opposed to any form of standardized test-based appraisal. We propose conducting the APEP operation every fourth year in the foreseeable future. The result, if implemented, will add a history of student accomplishments to our history of institutional program and policy changes.

In retrospect, there were several important decision points in this project once it was underway. The first was to go ahead with a testing program based on the definitions we had rather than develop new ones in the context within which we were working. The second was to capitalize on the campuswide interest in general education as a result of system mandates to embed APEP skills into the new program and concomitantly to delineate the relationship between skills and content. Finally, we took the opportunity to build an evaluation component into the new campus policy and related it to subsequent program building,

thus ensuring that the spirit of APEP goals and processes has been institutionalized.

Primarily the results will be used for evaluating and pointing out needed revisions to our general education program. This information will be provided to deans of schools seeking participation in the program in general education, so that they may be aware of strengths and weaknesses the testing program has identified. For the public we can provide assurance that a systematic program evaluation project is in place, that it has an impact, and that program changes will be made where weaknessess are identified. We will be satisfied with the program only when all of the freshmen/ senior differences are apparent and significant.

We do not plan to release the results of the testing either to celebrate accomplishments or to deplore weaknesses. This information is to be used internally for educational purposes; to monitor, to point out areas requiring change, and to evaluate the effectiveness of our decisions. "Outsiders" with a need to know specific results will be allowed access to such information. (Contact the director of the Office of Institutional Research.)

Written By

Patricia Brose and John G. Safarik

Institutional Contacts

Robin Wilson, president
Patricia Brose, professor of education
John G. Safarik, director of institutional research

North Adams State College (MA)

Institutional Profile

Located in North Adams, a city of 17,000 in the northwestern corner of Massachusetts, North Adams State College (NASC) is one of ten state colleges in the Massachusetts System of Public Higher Education. Chartered in 1894, it currently enrolls approximately 2,100 FTE daytime students mostly from Massachusetts but also representing fifteen states and five foreign countries. North Adams State College is primarily a residential college with approximately 80 percent of its students living on or near the campus.

The college offers fifteen majors and eleven minors, as well as a variety of special programs. Approximately 30 percent of the students are enrolled as business majors and 10 percent each in teacher preparation programs, sociology, and computer science. Other major programs are in biology, chemistry, English, history, interdisciplinary studies, mathematics, medical technology, philosophy, physics, and psychology.

Background

In the fall of 1977, North Adams State College instituted a new general education curriculum to replace a purely distributive one. The new curriculum established seven general education categories, defined the purpose of each category, and identified specific courses for each category. During the next two years, the curriculum committee monitored the program, made decisions about what new courses should be allowed into the curriculum, and undertook a limited evaluation of the program in the spring of 1979. There was general agreement that a more rigorous evaluation of the curriculum would be desirable.

It was in this context that the college first heard of the Academic Program Evaluation Project. James R. Roach, vice president for academic affairs, submitted a miniproposal to AASCU, and the college was notified in August 1979 that it would be one of ten colleges and universities participating in the project.

Implementation

Preliminary Phase

Several issues were paramount during this early period of the project. What would be the best way to evaluate the general education curriculum? How could an evaluation be carried out sc that the faculty would be most supportive? To what extent should the APEP guidelines drive the program evaluation? In general, there was agreement about the need to define each category of the general education curriculum more carefully, that broad faculty participation was very important in this process, and that the college should use the APEP guidelines as a resource rather than be driven by them.

At a faculty meeting in September 1979, it was announced that the college would be participating in the APEP. Few details were provided at that time, but generally the message was that the project would be of assistance in evaluating the new general education curriculum.

Following receipt of some start-up guidelines from the national office in mid-October, the campus director and coordinator of the project discussed alternative strategies for getting the project underway at NASC. The college's incentive awards committee, in cooperation with the curriculum committee, was planning a "Day of Conversation" to involve the faculty and administration in discussing and clarifying the meaning of general education. It was decided to capitalize on this event and to consider it the beginning activity of the project.

Approximately thirty-five of the total faculty of 100 attended this Day of Conversation. After a few opening remarks, faculty members formed interdisciplinary small groups to discuss such questions as: What is your definition of general education? What do you think it means to be an educated person? What intellectual abilities should a student be able to perform as a result of a good general education? What criteria should the curriculum committee be using when deciding whether a particular course should be accepted as a general education course?

Once groups had finished their discussion, group leaders took part in a panel presentation of the findings and fielded questions from the floor. The final segment of the conference was an address by the president, who spoke in favor of general education and identified a number of problems and possibilities related to general education. Reference was made to APEP and the role it could play in helping the college improve the general education curriculum.

The curriculum committee was given overall responsibility for overseeing the project on campus. The primary tasks for the first year were identified as follows:

- Elaborate on and clarify the objectives of each category of the general education curriculum.
- Define the objectives in terms that will allow faculty members to determine when they are being addressed.
- Conduct an assessment of existing general education courses to determine when the objectives are being addressed.
- Take steps designed to increase the extent to which the objectives are being addressed.
- Begin work on an evaluation design to be implemented during the second year of the project.

In discussing their charge, committee members agreed that it probably would be wise to downplay any mention of a general evaluation effort during the early part of the project and to speak in terms of the project's helping the faculty do a better job of helping students achieve the objectives of general education. It was decided not to take any further steps with the project before the national meeting of all project participants, scheduled for December 1979.

Conceptual Definitions

The national meeting marked the first turning point for the project at NASC. Prior to attending the conference, the general plan had been to make more specific the objectives of each general education category as a foundation for subsequent assessment and evaluation and for APEP to constitute one part of this larger effort. After the Washington meeting, the broader effort gradually was postponed and attention focused on defining more clearly the general education abilities of communication, quantification, analysis, and synthesis. .

The decision to concentrate on these intellectual abilities rather than attempting to define specifically all the concepts and abilities associated with the various general education categories was made for several reasons. Simple time constraints were a primary consideration. Also, there was a feeling that the college should fulfill its obligations to the national project. A strategy was needed that would allow focusing on the four intellectual abilities identified by APEP without losing sight of broader local objectives. Also, the selected strategy would have to ensure work on conceptual definitions did not become an arid intellectual exercise and involve a significant number of the faculty.

Four members of the faculty were commissioned to write working papers on the four intellectual abilities of communication, quantification, analysis, and synthesis. These working papers were to examine literature on the abilities and to describe the various ways the abilities are defined. All members of the faculty and administration received a memorandum summarizing the background of the project, describing progress to date, and inviting them to attend a meeting to discuss the four papers.

Over a quarter of the faculty attended this meeting. The authors of the papers made brief presentations, interdisciplinary subgroups discussed the papers, and group leaders subsequently reported on reactions to the papers. Revised papers were distributed to the entire faculty for review and comment. The revised papers described the conceptual territory for each of the four intellectual abilities but did not attempt concise definitions. Was this sufficient as a foundation for faculty members to work on operational definitions within their own disciplines? Would anything be gained by abstracting from the richness of the papers a concise conceptual definition and then having to move back to more concrete behavioral statements?

The curriculum committee decided against formulating tight conceptual definitions. Instead, outlines of the four papers were taken to Washington and guidance sought on how best to proceed. At the Washington meeting it became clear that some institutions had prepared tight conceptual definitions, others had listed subcomponents of the intellectual abilities, and other institutions had done both. It became clear that North Adams would need to move quickly to formulate operational definitions from its conceptual descriptions.

Feeling the pressure of time, not wishing to lose momentum on the project, and desiring to involve the faculty in operations as soon as possible, the project committee decided to draw from the list of subcomponents that had been developed by other participating institutions those that were most clearly stated. Thus, North Adams State moved from its four papers describing the rich conceptual territory associated with each of the intellectual abilities to lists of subcomponents for each of those abilities. These were to be the reference points for operation.

Operational Definitions

The need to move from conceptual discussions to operational definitions was clearly recognized. Faculty members would not be likely to sustain their involvement unless they saw some direct implications of the project for their own work in the classroom. At the same time, it was important to keep in mind the purpose of moving toward an evaluation of the general

education curriculum. What strategy would be most effective?

One possibility was to bring together faculty members teaching courses in the various general education categories, group them by categories, and ask them to work on operational definitions for communication, analysis, synthesis, and quantification. Given such definitions, an instrument could be devised or selected for evaluating student performance across the entire program.

The turning point for the entire project at North Adams State came when it was decided that the best approach to an eventual evaluation of the general education curriculum was to begin by having individual faculty members work within the context of their own courses to improve student performance of the intellectual abilities. The rationale was that if faculty members were to get involved in a systematic assessment of student performance within their own courses they would be more receptive to a systematic evaluation of the entire general education program. The project would later include some beginning efforts at programwide evaluation but primary emphasis would be given to work at the level of courses.

With this strategy in mind, a memorandum was sent to all faculty members inviting them to make a commitment to the remainder of the project and spelling out what would be involved. Each participating faculty member would focus on a course he or she taught and, in conference settings, work with three to five others in support groups to:

- share the details of their courses as currently taught
- identify specific ways students might communicate, quantify, analyze, and synthesize
- explore ways to foster more effective intellectual functioning in the content of each course
- design methods for evaluating student performance of the intellectual abilities.

Over twenty faculty members from most of the disciplines volunteered to take part.

Development of Assessment Tools

The APEP paradigm developed by consultants to the project provided a common frame of reference for participating institutions. Although North Adams State had chosen to work first within the context of individual courses, it was important that work proceed with reference to the paradigm. This meant that participating faculty members would need to develop assessment items and criteria for their own classes, assess student performance and factors affecting that performance, make judgments about student performance and their course, and decide on possible changes.

Accordingly, two meetings were held with volunteer faculty members in May 1980 and a two-day retreat followed in June. The first meeting was viewed as an orientation session. Each participant was provided with a notebook containing key documents from the central office and memoranda and papers that had been developed locally. Expectations and ground rules for the project were made clear and questions answered. Throughout this first meeting, the emphasis was on collegiality.

The first working session was held in late May 1980, and the objective was simply to get faculty members talking about their own courses and ways that they have sought to challenge students to communicate, analyze, synthesize, and quantify. Each participant brought copies of materials for the course he or she had selected and, in interdisciplinary small groups, shared with others the objectives for the course, instructional strategies commonly used, and evaluation modes. Later in the session, each group focused on one of the intellectual abilities and discussed specific ways they sought to foster that ability within their several courses. By the end of this first meeting, faculty members had become comfortable talking with one another about their courses and about ways of fostering higher-level thinking. They especially felt good about talking with faculty members from other disciplines and expressed a willingness to participate in the two-day retreat planned for mid-June.

This retreat was seen as critical to the future success of the project. It would represent the first time faculty members would be asked to look analytically at their own practices and to consider ways of writing more effective assessment items. Prior to the retreat, all participants were sent a "Homework Packet" asking them to prepare an outline of their course and, on a separate sheet containing all the subskills of the several intellectual abilities, to indicate whether each subskill was one they would like to emphasize more explicitly or whether it seemed irrelevant to the course. The intention was for each faculty member to arrive at the conference having thought about the various intellectual abilities in detail.

The basic strategy for this workshop was to engage the participants in the writing of assessment items geared to a subcomponent of a selected intellectual ability and to discuss criteria for judging student performance on the item. Each subgroup identified a single component of an intellectual ability, determined some possible activities that would call on students to perform that subskill, and then designed an actual assignment (complete with explicit instructions) to be shared later with all participants.

The next morning, each subgroup continued to focus on the assignment they had designed but this time attention was directed to anticipating the range of performances and on developing criteria they would use to discriminate among these performances. In the process, several groups found it necessary to rework the wording of their assignment in order to make it clearer.

These exercises in operationalizing were interspersed with discussion sessions. For example, many found it difficult to concentrate on a single component of an intellectual ability but felt less uneasy when it was acknowledged that probably many intellectual operations would be involved in any given assignment and that they were being asked simply to give most atttention to a particular component. Large-group discussions of such issues and the informal socializing that occurred during the cocktail hour, at meal times, and the extended nightcap session all contributed to the growing identity of the group. Again and again, participants acknowledged the value of having a chance to talk about teaching and learning with those representing other disciplines.

Everything that had been done during the first year of the project can be seen as being preliminary to going operational. Conceptual descriptions had been written, subcomponents identified, and the faculty acquainted with what is involved in preparing assessment items and criteria. During the next year of the project, each faculty members would build on this foundation, applying what they had learned to one of their own courses.

To sustain momentum, participants were sent a summary of the conference, a draft plan of activities for the next year, and a resource packet to aid in their individual planning. Each participant was encouraged to establish clearly the objectives for the course they had selected, concentrate on one or a few components of an intellectual ability, and be imaginative in designing the assessment items. An overview of different assessment modes and their advantages and disadvantages was provided. Finally, suggestions were given that would assist the faculty in specifying criteria for judging student performance levels.

Throughout the 1980-1981 academic year, participants tried out their assessment items, refined them, and tried them again. This process was facilitated by occasional meetings with other members of their support groups, by a progress-sharing conference in October, and by a mid-year conference. Most faculty members chose to concentrate primarily on one or another subcomponent of analysis, whereas others concentrated on communication, synthesis, or valuing. Some faculty members were very systematic in their experimentation; others found it difficult to keep their focus on a particular intellectual subcomponent. Again, a real benefit of involvement was the interaction among representatives of different disciplines.

The APEP paradigm being developed and refined during that year called not only for the preparation of assessment items and criteria for judging performance but also for faculty members to consider factors that might affect student performance. Informally, participants had discussed such factors early in the project, when they shared with one another ways they sought to foster higher-level thinking. This stage asked them to be more rigorous in identifying such factors and in relating them to the performance on assessment items they had designed.

During the two-day conference in January 1981, small groups consisting of an interdisciplinary mix of faculty members were asked to list teacher and student behaviors they felt helped students analyze, communicate, quantify, and synthesize. Each group concentrated on a particular intellectual ability and the list of items was then shared with all participants.

With this discussion as background, each faculty member was asked to think about his or her course in detail and to indicate the means they were using to evaluate student performance (essays, journal, class participation, etc.), to list any of their behaviors that might foster intellectual ability, and to identify student behaviors that might affect their performance.

By the end of the 1980-1981 academic year, each faculty member had designed and refined assessment items for his or her course, considered criteria for judging student performance on those items, and identified teacher and student behaviors that might affect performance.

Program Assessment

While much of the 1980-1981 academic year was devoted to faculty members working within the context of their own courses, steps also were being taken to move toward an evaluation of the entire general education program. At a meeting of participants in October 1980, sample assessment items from the Educational Testing Services' measures for general education and from McBer and Company's *Analysis of Argument Test* were duplicated for examination. Faculty members answered each of the items and were then asked to consider how well the items measured analysis, communication, quantification, or synthesis as defined by the subcomponents identified the previous spring.

Reactions were largely negative. Most of the discussion focused on the ETS items. There were differences of opinion on the "correct" answers. The wording of the questions was criticized. Many expressed doubt that multiple-choice items were appropriate for measuring high-level intellectual abilities.

In spite of these reactions, participants were asked to administer at least one of the ETS tests to their classes prior to the end of the semester. It was suggested that they might find it interesting to correlate students' results on whichever test they selected with students' performance on local assessment items they had designed. Approximately half of the participants agreed to administer at least one of the ETS tests. The results were scored and returned but for a variety of reasons faculty members did not explore correlations.

Paralleling the experimentation with ETS items was exploration of the instruments developed by the ACT's College Outcome Measures Project. All members of the steering committee attended a two-day workshop in Albany, New York, and came away impressed with what they had experienced. At this point of the project, however, the primary emphasis was still on faculty members developing course-specific assessment items, so further consideration of ACT/COMP tests was postponed.

A further step to gather programwide information was undertaken during the January 1981 two-day conference described earlier. Participants selected four items each for analysis, synthesis, communication, and quantification from the most reliable forms of the ETS measures for general education. The resulting sixteen items were assembled into a composite test for measuring high-level thinking abilities. This test was administered to 482 students in February 1981. Later in the spring, 338 of these same students took another form of the examination, thus providing the college with at least some information on student performance in all majors and classes.

To gather some program-portrayal data, it was decided to try to identify the SAT scores for each student taking the composite examinations and to ask faculty members to include the following two items on a questionnaire for student evaluations of courses and instructors:

- "This course required me to think, not just memorize."
- "The instructor provided constructive written or oral comments on students' work."

The results of these programwide efforts, together with the data gathered on course-specific items and program portrayal factors, constituted North Adams State College's work on the latter stages of the APEP paradigm.

Outcomes

Because of North Adams State's two-pronged approach, data were generated at both the course and program levels. From the beginning of the project, faculty participants had been assured they would maintain control of any data generated by their course-specific assessment items and that they would not be required to share results with the community at large. However, they were invited to write individual case histories of their experience with the project and to reflect on changes that might be in order. A majority of the participants did prepare case histories, some more detailed than others. All of them, however, commented favorably about their project experience. Most described the kinds of assessment items they had designed and the procedures they had used for stimulating higher-level thinking among students. Some of the respondents commented on how the project might be continued; all of them felt one of the primary benefits had been the involvement with faculty members from other disciplines.

At the program level, data were generated primarily by the two composite tests administered in spring 1981. In general, the results of both tests revealed clearly that seniors performed better than juniors, juniors better than sophomores, and sophomores better than freshmen. The following tables show the percentage of correct responses for each segment of the composite tests:

First Test	*Number*	*Analysis*	*Synthesis*	*Mathematics*	*Communication*	*Overall*
Class '81	78	58	78	71	69	67
Class '82	110	53	76	65	54	61
Class '83	104	45	70	60	55	56
Class '84	190	43	61	57	41	49

Second Test	*Number*	*Analysis*	*Synthesis*	*Mathematics*	*Communication*	*Overall*
Class '81	60	63	51	63	61	57
Class '82	75	67	43	66	53	55
Class '83	75	54	44	61	47	49
Class '84	128	52	40	59	38	44

It was commonly agreed that the items on the second composite test were either more difficult or less clear than those on the first test. Also, many of the students expressed their impression that they "had already taken" this test and therefore had not taken it seriously.

Unfortunately, neither of the attempts to gather programwide information on factors that might affect student performance were successful. SAT information on students taking the composite tests was not readily available, and it was decided not to carry out a record-by-record search. Second, only a few faculty members chose to include on the Student Instructional Rating Form the two questions designed to elicit student perceptions of the challenge of the course and the amount of instructor feedback.

To summarize, the only hard data generated were the results of the composite examinations. This is not surprising in light of the college's chosen strategy, which was to lay the foundation for collegewide acceptance of a more systematic evaluation of the general education curriculum. By minimizing the emphasis on "evaluation" and putting primary emphasis on getting faculty members involved in assessment activities, the hope was that subsequent programwide evaluation efforts would reap greater long-range benefits than had they been initiated early on.

Conclusions

Policy Development and Decision Making

What have been some of the policy development and decision-making outcomes of North Adams State involvement in the APEP project? Considering the program level, one concrete result was the revision in spring 1981 of the preamble and category descriptions for the general education curriculum. Because of the greater awareness on campus of the importance of higher-level thinking, the preamble makes reference to the importance of reasoning ability, and the "capstone" category of the curriculum calls for students to demonstrate the intellectual abilities of communicating, analyzing, synthesizing, and valuing.

At the time this case history was written, the curriculum committee was beginning a thorough review of the general education curriculum that would include clarifying criteria for reviewing and screening general education courses and establishing more rigorous procedures for determining which courses should be part of the curriculum. Among the criteria almost certain to be applied would be the extent to which a given course is designed to foster at least one of the intellectual abilities identified by APEP.

These developments can be fairly described as being a natural outgrowth of the college's participation in the project. When a substantial number of faculty members became involved and were challenged to think about better ways to foster higher level thinking in their courses, it became a logical extension of their activities to institutionalize a more explicit set of policies having as a common goal better communicating, analyzing, synthesizing, quantifying, and valuing.

Of course, much remains to be done. It is one thing to establish policies, quite another to implement them. Given a solid core of faculty support and the experience gained during the life of the project, successful implementation certainly is more probable than would otherwise have been the case.

Some Reflections and Recommendations

In thinking back on the activities, strategies, and decisions made over the 2-1/2 years of the project, several reflections and recommendations come to mind. At the beginning of the project, it might have been best for the college to proceed with the original plan to have groups of faculty members seek to define more clearly the various intellectual abilities implicit within the category descriptions of the general education curriculum. Very likely, these would have included those identified by APEP, but it would have been better for the faculty to "discover" this than to have them imported from outside.

The preparation of four position papers during January 1980 was a good idea as was the follow-up conference at which they were discussed. In restrospect, it would have been better to work on identifying the components of the intellectual abilities rather than stopping with the descriptive papers. The price paid was that the faculty never had full ownership of the components subsequently selected from among those developed at other institutions.

The strategy of first working with faculty members at the course level was a sound decision and constitutes the major recommendation to be made to institutions of comparable size and characteristics. Faculty members responded enthusiastically and followed through with no reward other than their own feelings of professional growth. Eventual policy decisions were an outgrowth of the entire process rather than the result of a comprehensive, programwide evaluation that might not have been so well received.

Part of the strategy developed in April 1980 called for faculty members to work in small support groups consisting of three to five members convened by a member of the steering committee. Such groups are recommended strongly, even though North Adams State was unable to implement them fully. Steering committee members simply did not have enough time to assimilate all of the project literature and to work with their groups. Should other institutions about this approach, it is recommended that they try to provide some release time for group leaders.

It is extremely important for an institution to be aware of its own faculty's needs and motivations and to use the APEP guidelines as a reference tool rather than as a recipe. The steps of the guidelines are helpful in strengthening the process of program evaluation but should not be employed without constant reference to the many forces at work within an institution.

Perhaps the greatest payoff of North Adams State's involvement in the APEP project was the contribution it made to faculty morale. Again and again, faculty members spoke of the value of meeting with others to talk about teaching and learning.

Written By

Richard E. Markham

Institutional Contact

James R. Roach, vice president for academic affairs
Richard E. Markham, chairman, Department of Interdisciplinary Studies

Ramapo College of New Jersey

Institutional Profile

Ramapo College, located twenty-five miles northwest of New York City, is one of nine state colleges is New Jersey. As it began its thirteenth year in September 1983, over one-third on its 4,500 students were part timers and more than 80 percent were commuters. Approximately 25 percent of the students were over age twenty-three.

Twenty-one major programs leading to the B.A., B.S., and B.S.W. degrees are offered by approximately 135 full-time faculty members, 90 percent of whom hold the doctorate or equivalent terminal degrees, and thirty equivalent part-time faculty members. The faculty has organized itself into six degree-certifying schools, structured nondepartmentally and oriented by academic philosophy toward interdisciplinary and theme-oriented studies, administration, and business, American and international studies, contemporary arts, environmental studies, social science and human services, and theoretical and applied science. In addition, there are two academic divisions: basic studies and physical education. (As of 1985, the School of Environmental Studies was merged with the School of Theoretical and Applied Science and the Division of Physical Education was disbanded.)

Background

Although as a "second-generation" state college Ramapo had the opportunity to develop a faculty organization and culture particularly shaped by its interdisciplinary orientation, its problems and concerns were not unlike those of the other public state colleges and universities associated with the Academic Program Evaluation Project. Like most colleges over the past decade, Ramapo had endured a prolonged controversy over the nature and purpose of general education. Three elements had been especially prominent in the Ramapo discussions: the integration of critical method and subject content in disciplinary instruction; the need for demonstrated proficiency in reading, writing, and quantification; and the need for a vertical (sequential) coherence and horizontal (synchronic) coherence in courses. The decision to submit a proposal to AASCU was prompted, to a large degree, by the good match between issues prevalent in the Ramapo context and APEP's overall interests in the primary intellectual or generic competencies of communication, critical thinking, and quantification.

Other promptings, of a more external nature, occurred with the introduction in 1978 of a statewide test for all entering public college students, the *New Jersey College Basic Skills Placement Test.* This test drew public attention to "basic skills" and accelerated the development of remedial and developmental programs at all New Jersey public colleges. At Ramapo the growth of these programs increased faculty sophistication about the testing, assessment, and teaching of underprepared students. It was agreed by all that basic skills needed to be improved, but that was only the first stage of student development toward intellectual abilities commensurate with the baccalaureate.

In addition to APEP's emphasis on inteliectual skills as a critical element in the baccalaureate, APEP's purposes and goals seemed, in 1979, equally timely and pertinent to Ramapo's record of institutional growth. Approaching the end of its first decade of operation, the college had experienced in its early years rapid enrollment growth and a burgeoning curriculum. The formation and later reformation of the Ramapo schools had by 1979 clarified the college's instructional content and mission. This exercise of curriculum building by a comparatively young and mostly tenured faculty was further taxed by processes that brought the majority of the faculty to tenure by 1979. During this first phase, self-study for the purposes of accreditation and certification, in addition to the cycles of a major program evaluation, also commanded the faculty's attention.

Given Ramapo's organizational and curricular growth and the great diversity of its student body, it was time to examine how the curriculum contributed to the attainment by Ramapo students of the intellectual skills of communication, critical thinking, and quantification. Given the past attention to self-study, such an examination, it was anticipated, would happily be done in concert with nine other state colleges and universities. APEP therefore became an opportunity for Ramapo to conduct a comparative

examination. Viewing its own purposes and outcomes within the context of nationally representative state colleges and universities would place Ramapo's own efforts in an important perspective.

Implementation

Although faculty members and students will, at an appropriate distance, view the purposes of such a project as APEP as useful and salutatory, bringing APEP directly into the curriculum and classes was another matter. Other chapters by APEP participant institutions have made sufficient reference to examples of faculty resistance to assessment and measurement and the causes for that resistance. The universality of this situation only accentuated its importance. The challenge of how to determine the "ownership" of APEP and how to make it a legitimate institutional activity is answered largely by local governance traditions, past practices of institutional research, and the faculty culture. It soon became evident to all the APEP institutions that to avoid the suggestion of APEP as simply administrative oversight, the pathway to answering the difficult and ambiguous questions raised by APEP had to begin with and continously pass through the faculty. The APEP goals needed to be turned into intellectual questions formulated by faculty members and shaped into techniques of inquiry and measurement that would satisfy the protocols of study and research. If APEP were to have any relation to the study of teaching and learning in a faculty culture, that inquiry had to be formulated and executed by a representative faculty group.

The choice of a representative group at Ramapo followed the same lines of governance and consultation that had proven durable in the rapid development of the curriculum. Participants were invited and confirmed by the faculty of the schools and divisions. The school remained the locus of faculty governance, with all-college matters treated in standing committees comprising representative members elected by the councils of the academic units. The sturdiness of this unit governance process permitted the emergence of an eleven-person committee, which remained remarkably stable in membership during the two and more years of its activities.[1] Such committees are not simply elected. They cohere only after many hours of informal mettings, scheduled presentations at unit meetings, including the circulation of formal summary material, and passage through proven channels of faculty governance. Such a process led, at Ramapo, to the threshold of what the APEP consultants liked to call "ownership."

First-Phase APEP Efforts

By spring of 1980, one year after Ramapo formally joined the consortium of APEP institutions, the faculty committee on APEP had defined its membership and moved closer to understanding how its collective talents could be brought to bear on the intellectual and methodological issues raised by the purposes of APEP and the relationship of APEP's goals to the curricular intentions and outcomes of Ramapo's curriculum. In a summer 1980 interim report the committee summarized its initial purposes in the development of APEP and forecasted its activities for 1980-81:

- to acquaint the Ramapo community with the purposes of the project and to solicit and encourage the participation of the faculty and staff
- to review the pertinent critical literature on the definition and assessment of undergraduate college outcomes, particularly as they address the broad intellectual skill areas of communication, critical thinking, and quantification
- to examine the structure and kind of Ramapo's undergraduate instruction and to confer with the academic units about the relationship of that instruction to the project's purposes
- to propose a plan for an evaluation design to carry out the APEP goals at Ramapo during the 1980-81 academic year.

In 1979-80 the committee devoted most of its efforts to the first task of acquainting the Ramapo community with the purposes of APEP. For more than two months the committee met regularly and, with the assistance of the library staff, Ramapo's Office of Institutional Research, and consultants from within the Ramapo faculty and from the outside, pursued its inquiry into the purposes outlined above.

Definitions of Intellectual Skills

One striking phenomenon observed by the committee early on was that definitions of generic intellectual competencies are quite primitive: literature searches revealed little and the solicitation of consultants or experts resulted in few individuals who had a sophisticated knowledge of the area. The committee did find papers distributed by the AASCU Resource Center for Planned Change staff quite useful in this regard, especially Gary Woditsch's paper on generic competency

and Jonathan Warren's survey of how California professors characterized their best students.

The APEP member institutions were initially invited to determine baccalaureate-level intellectual skills using the categories of communication, analysis, synthesis, and quantification. In early discussions on the nature of these intellectual abilities, committee members realized they were subject to a number of limitations of different kinds. For the purposes of this study the skills were to be exit-level rather than entry-level remedial. Ideally, they should be, in Gary Woditsch's language, "recursive and ubiquitous generic competencies." In practice, however, these abilities could not be fully abstracted from content and instructional context. The committee believed these intellectual abilities overlapped more than they were separable. In fact, it was unable to make any useful distinction between analysis and synthesis and decided that the two would best be described as "critical thinking."

Furthermore, abilities could not be merely operational or mechanical; critical abilities such as the recognition of problems were essential elements. These abilities were not to be tied into any major; they were to be generic. It was recognized that the instructional intentions of each school might emphasize different aspects of these abilities. For example, the School of Contemporary Arts might emphasize proportion and other less numerical relations in quantification than would, let us say, the School of Theoretical and Applied Science. The committee decided that its inquiry should be concerned with the student's capacity to demonstrate these intellectual abilities, not merely whether he or she had the opportunity to learn them. Although practical problems might arise, the definitions of these abilities should not, at least initially, be limited by what the faculty could teach. Finally, the committee acknowledged that its wish to define would be constrained by the economic and time limits of what is testable.

For a time the committee thought it best to talk about a constellation of elements in these skills, since the intersections of the three (communication, quantification, and critical thinking) were more common than the unique features belonging to any one of them. Critical thinking plays an essential role in the processes of quantification, and in communication one must communicate, at least to oneself, the self correctives of logic. Modeling in quantification draws extensively from the logic of analogy. Probability and statistics are frequently displayed as visual communication. Perhaps, the committee speculated, it is only the more operational features of any of the competencies that one might consider exclusive to each: computer utilization and calculation in quantification, the proper use of grammar in communication.

Baccalaureate-Level Generic Competencies

Within these prescriptions, the planning committee devised the following descriptions:

- *Communication.* This generic skill is something more than a giving and receiving of messages; it includes a number of elements such as the capacity to listen and speak well, abilities that presently exceed practical limitations of campuswide testing. The committee recognized that communication as a generic ability includes writing, reading, and visual and historical communication. Writing includes an operational knowledge of fundamentals of composition and the grammar of standard American English, sufficient vocabulary, capacity to make a thesis structurally evident in prose, and consistency of presentation. In reading, one must be able to identify the theme of a written piece of material and to distinguish between fact and opinion, literate and intentional meanings, and the connotation and denotation of terms. One must, furthermore, be able to recognize symbols and have a sufficient expressive and analytical vocabulary. Visual communication consists of the interpretation of the formal structures of visual presentations, such as graphs or works or art, and the recognition that these structures emerge from the creative interplay between materials and concept. Historical communication is characterized by its application of models taken from the sciences (e.g., forces) to historical phenomena and its elucidation of the historical conditioning of the meanings of terms.
- *Critical Thinking.* Analysis and synthesis, as has been indicated, were combined as critical thinking. As essential elements of critical thinking, problem detection, argument analysis, premise identification, judgment of logical validity, and detection of alternate premises proved more testable than critical attitudes, position taking, recognition of systematic and extrasystematic limitations on methods of thought, the use of models, and elaboration of context.
- *Quantification.* The committee recognized that the competency of quantification is a process that includes more than mathematics as it would be taught sequentially in school or college. Rather, it is processual: the application of mathematically related skills and concepts to various types of applied problem solving. These mathematical skills include counting and arithmetic, measurement, determination of

geometric relationships, and algebraic, trigonometric, and statistical manipulations. These are applied in many ways, including modeling techniques and various problem solving work in the sciences and social sciences.

Evaluation Design

Despite the realization that generic competencies should suit the realities of assessment, the definitions of these competencies still proved in practice to be too lofty. Practical constraints of resources and time, and the limits of what the chosen tests purported to measure, made it necessary to draw even more restrictive lines around them to make student progress in these competencies more accessible. Further, the committee thought it excessively bold to attempt to draw up tests to measure achievement of such complex intellectual characteristics, especially when no useful validation of the additional test elements would have been available. Thus, the faculty committee felt constrained generally to use instruments already available commercially or through AASCU consultants and correlate test results with transcripts and other data already available. It was hoped that a few of the nonmeasurable features of the skills could be analyzed through in-depth interviewing of faculty members and students, coupled with a careful examination of core course syllabi and final examinations.

The committee, together with a social psychologist from the School of Theoretical and Applied Science, acting as consultant, designed in summer 1980 an evaluation of student progress in the competencies. The goal of the evaluation was to provide information on which specific combination of learner characteristics significantly predicts adeptness in communication, quantification, and critical thinking as measured by the testing instruments. These learner characteristics were listed as age, sex, academic year (freshmen through senior), SAT scores, total credits completed, and number of general education courses completed.

It was recognized that a longitudinal study was not possible within the time frame of APEP. Hence it was decided to carry on a cross-sectional study for the academic year 1980-81 with some hope that longitudinal follow-up might be possible within an extension of APEP. For the purposes of measuring written communication, a locally developed, thirty-minute essay test was devised, to be rated through two independent scores by faculty readers. This test format, as well as the rating procedure, was chosen because it reflected the essay format and holistic scoring procedures used in the essay component of the *New Jersey College Basic Skills Placement Test.* To measure critical thinking and quantification a number of comercially available instruments were examined. After pilot testing the Watson-Glaser Critical Thinking Appraisal test and the Sequential Test of Educational Progress (STEP) mathematics test with a subsample of students in August 1980, the committee decided to use these instruments. It was recognized that the student sample would have to consist of all volunteers, optimally 500 or more. A release form was drawn up to request permission for use of student's records and to match data on test scores with student's social security numbers.

Data from learner characteristics would serve useful predictive functions. Moreover, it was felt that age, sex, and attitudes would provide useful profiles of the subjects; academic year, scores from the SAT, the Basic Skills Tests would permit measurement of value added or lost. Also, the intervalidation, or the lack of it, between scores and grade-point average would be useful.

Measures in the Classroom

Once the measures had been selected, the committee faced the task of how to obtain the cooperation and participation of the students who would need to complete, in a reasonably consistent testing environment, the measures chosen. Should a student sample be recruited by letter of invitation, which may or may not include the promise of remuneration or other "incentives," or should they be tested during a scheduled class period with the understanding, of course, that their participation was voluntary? The committee decided to use scheduled class periods in the first week of the semester. The committee concentrated on those classes in which students from Ramapo's three selected schools would be enrolled and during enough hours of the day to ensure that adult part-time students would be included in the sample. To ensure a sufficient sample of senior students, the committee included as many senior seminar courses as possible. The decision to use the opening sessions of a class was motivated by the committee's belief that it would become increasingly difficult to obtain the cooperation of the faculty once course content had been fully committed. Student attendance is usually highest during first sessions, and it would still be early enough in the course for the students not to regard testing efforts as an intrusion on their attention.

Outcomes

Though the testing process was fraught with logistical problems, there were great rewards. Once a class was identified and the permission of the faculty obtained, an APEP committee member or another fully briefed faculty or staff member would appear at the class. The overall purposes of the APEP research project were outlined, and the students were assured that anonymity would be preserved and that the outcomes of the testing would be reported to any student who wished a follow-up interview. Massive defection did not materialize. In some instances individual students declined to participate and in one isolated (but still notorious) case an entire class got up and left, prompted by a faculty member who had orginally declared his interest in cooperating.

Generally, the committee found itself involved in animated discussions about the relationship between testing and learning and the function of educationl research. It appeared that students would give their time when they believed that the requested tests would contribute to institutional self-understanding and therefore to greater understanding of the intellectual abilities of the students. Long after the test period students and faculty members would return to the issues raised during the test periods, and APEP's purposes would enter into discussions of both an informal and formal nature. The testing periods gave APEP a pervasive status in the community.

Interpreting the Data

The two stages of APEP at Ramapo could be largely described as, first, a process of informing the academic community of the purposes of APEP and, second, a faculty-directed inquiry leading to the design and execution of an educational research study. A detailed summary of method and results was prepared and submitted by Ramapo to AASCU and FIPSE as the technical report in our final case study.[2] It would not be useful to replicate, in this context, the details of that report; however, an answer to one basic question can be given.

Does the baccalaureate program enhance student's abilities in the generic competencies of communication, critical thinking, and quantification as measured by these instruments?

The correlation coefficients indicated that a higher count of credits was associated with higher scores on the indices of cognitive skills. However, it must be immediately added that more advanced students have on the average more of the abilities necessary to survive in an academic program than those who have fewer credits. When test data were corrected for such factors, it seems that the program did not enhance (or diminish) the "cognitive competencies" as measured. In order to isolate the variables used for correlation with the measures, a series of regression analyses were conducted on test scores in relation to credit count, age, grade-point average, sex, Ramapo school of enrollment, SAT verbal and SAT math scores.

The regression analyses suggested that in quantification higher credits, higher grade-point average, school of enrollment, and SAT math proved to be significant predictors of STEP mathematics testing scores. Age and sex were or lesser significance: "adult" students did not perform any better of this measure than did "traditional" students.

Some measurable progress in critical thinking through the baccalaureate program was discernible. This factor was isolatable in a statistically significant way, even though grade-point average and school of enrollment proved to be significant predictors, with the verbal score the more significant or the two. Age and sex, as with the STEP test, were not noteworthy predictors.

Thus educators can take some comfort in knowing that whatever is measured on the Watson-Glaser test is enhanced to some extent by exposure to college studies. But data from the test also seemed to indicate that students with greater reading speed and comprehension did better than those less able. Whether the results are measures of reading ability or thinking ability or some generic competency from which both flow is a vexing question, which indicates the difficulty in assigning cause and description to cognitive skills.

For whatever the local essay test measured in communication ability, there was no noteworthy difference between the average scores obtained by freshmen and seniors. Grade-point average and the SAT verbal score revealed themselves as significant predictors.

Comm. .should be made here about the dominance of the SAT scores in all of our analyses of student competencies. For the past decade higher education has moved to improve equality of access, but the results at Ramapo suggest that we have not fared as well with "equality of competence." Academic aptitude and preparedness that a student brings to college, as suggested by the SAT scores, remain telling predictors of the level of competency in generic skills at graduation. The baccalaureate experience does not appear to compensate and "equalize" the different levels of preparedness revealed at entry to college. Low-achieving students, as predicted by SAT, remain

low-achieving students on the indices used for this study.

Conclusions

Many questions could be raised in response to the above outlined results and their interpretation. Faculty members who work with students over the span of a baccalaureate program are struck by the evident progress made by certain students who do not initially manifest academic abilities. Their sense of themselves strengthens and the maturity of their perceptions enables them to display greater analytical and expressive abilities. Such "evidence" of progress suggests again that in the measures used only a small range of abilities were identified and that, although the narrowly psychometric intentions of the measures and their interpretations have been satisfied, the constellation of variables that contribute to intellectual growth remains larger and more elusive than these measures could grasp. In the case of the local essay used to assess communication, for example, there is now the strong suspicion that the format and scoring of that essay are insufficient to address the complexity of the competencies embodied in communication. Whereas a thirty-minute essay on a general topic rated on a four-point scale may be useful for initial placement purposes, such a measure is inadequate to address the scope and range of abilities that can be manifested in writing, no less in the comprehensive competency of communication.

It was the committee's judgment at all stages of its inquiry that much remains to be done in the development of measures of college outcomes that have wide acceptance and credibility among the academic community. It may well be that cognitive competencies ascend an asymptotic curve in which enormous progress is made in one's childhood but abilities level off in later years. Is the function of the college curriculum, then, to provide more and more content for abilities already well-established? Or can colleges do more to receive a reasonable return in values added to cognitive competencies?

Policy Directions: A Curricular Model for the Future

APEP has in part contributed to a more cooperative institutional research environment in which faculty members are no longer so suspicious about assessment as an intrusion into the progress of their students and have some limited faith in psychometric techniques. More important, however, it has provided a base of data from which longitudinal studies of the college's effectiveness in the teaching of generic competencies can be assessed. Although no specific decisions have been made about the thrust and method (or--importantly--*funding*) of such a study, it is likely that the research would be more limited, concentrating on the effectiveness of selected general education courses in developing generic competencies.

Programs in writing and elementary mathematics have for some time now been well underway. It is to be hoped that continued emphasis at Ramapo on these programs will produce some noticeable improvement in its students' competencies in these areas. There has been increased attention to peer tutoring in composition and an emphasis on problem solving and mathematical models in instruction in the basic studies division. Much remains to be done: writing should be required and carefully corrected in more classes in the disciplines. Still to be devised is a satisfactory and economically manageable definition of exit-level competency in quantification.

One might immediately object that the students already engage in much critical thinking in the disciplines. And there is no doubt that they do. But it is left to chance that the skills learned in the disciplines become generic competencies: it is left up to the student's own devices to incorporate that skill at a higher, more complex level so that it becomes ubiquitous or tranferable to other areas.

A New Program

The college is considering a program that addresses thinking as such. A sequence in instruction in thinking would consist of two stages: a one-semester course at the beginning of a student's college career, and a second semester somewhere toward the middle of the student's four years. The first component concentrates on the behaviors universally present among those who do well on IQ tests, with some introduction to the rules governing correct thinking. The second component would directly address the canons themselves.

Most students who have been admitted to college have by that fact alone demonstrated a persistence and raw intelligence that can serve as the stuff for future development in such a course. The first component of the curriculum is meant to free up those abilities, to unblock those behaviors which inhibit the flow of such a capacity. Or conversely it is meant to inculcate those behaviors universally characteristic of high scorers on IQ tests: selective attention, sustained analysis, analogizing, suspension of closure, and autocensorship. Students who have

already demonstrated this self-command by virtue of high scores on the SAT verbal test should be exempted, of course, because such work would be redundant. Such instruction does not address the actual forms of reasoning; the student would receive an introduction to precision in the use of language, an elementary use of the logic of necessary and sufficient conditions, some practice in elementary forms of argument, and some attention to models.

It is in the second component of the sequence that the methods of thought are addressed in their essentials: more complex uses of connective terms and quantifiers; case studies of patterns of scientific inference, including a look at the logic underlying statistical reasoning; a study of the patterns of derivation, including the use of matrices to establish the interconnections among many variables; the study of particular models, such as systems; and practice in problem solving, using such techniques as the elimination of alternatives and working backward from a known solution to its givens.

The college could evaluate the effectiveness of this program by comparing the scores in original SATs to scores from GRE tests. What steps can Ramapo take reasonably to ensure that such a sequence in thinking would noticeably improve students' scores on the SAT-GRE? Or, to put it in more fundamental terms, what are the criteria that must be satisfied to have an instructional program of high quality in thinking?

Besides the presumed criteria of attentiveness to detail of deadlines and academic rigor, other criteria should be pointed out. Course sections should be small enough for discussion and adequate enough in number for students to take the courses when they need them; course content should be integrated horizontally with other remedial or discipline-oriented courses the student is taking at that time and vertically integrated with other core courses; in addition, faculty and student tutors must have the special training necessary to assist students in knowing how to think more clearly. But the most important criterion for quality is that the courses must have the proper instructional modalities. This instruction must have at least three elements: the students verbalize the processes of their own thought in trying to solve problems presented to them; they have the opportunity to compare their thinking processes with models of good thinking, perhaps available on tape or from peer tutors, and are encouraged to take notes on the differences; and they are provided with feedback sufficiently detailed to be corrective and yet encouraging. Numerous experiments have shown these modalities to be successful in dramatically increasing students' scores on IQ and aptitude tests.[3]

Notes

[1] The faculty members of the APEP committee selected by the six schools and two academic divisions were Gordon Bear (social psychology), Eugene Beyers (psychology), Ronald Brady (philosophy), John Robert Cassidy--project director (philosophy), Robert Christoper--campus coordinator (literature), David Freund (photography), Thomas Goss (history), William Makofske (physics), David Turnage (physical education), Warner Wada (painting), Gabriella Wepner (mathematics). Helen Rogers (academic affairs) was the committee's administrative secretary.

[2] A copy of the technical report is available on request from Robert Christopher, Assistant Vice President for Academic Affairs, Ramapo College of New Jersey, Mahwah, NJ 07430.

[3] See, for example, Whimby's presentations on work by Benjamin Bloom and Lois Broder at the University of Chicago from 1945 to 1950 and his own work at Berkeley in 1966.

Written By

John Robert Cassidy and Robert Christopher

Institutional Contacts

Robert Christopher, vice president for academic affairs
John Robert Cassidy, professor of philosophy

Southern Illinois University at Edwardsville

Institutional Profile

Southern Illinois University at Edwardsville (SIUE) is one of the two four-year institutions constituting the Southern Illinois University System. The campus was established in 1949 when courses were offered in the metropolitan East St. Louis area. In the fall of 1965 major academic operations were centralized on the new Edwardsville campus. SIUE is located on 2,600 acres of rolling lands and woods flanking the Mississippi River southwest of Edwardsville. Eight buildings presently compose the main academic core.

There are approximately 560 full-time faculty members, 75 percent of whom possess doctoral degrees. Although many are distinguished nationally and internationally through special talents, publications, and research, teaching is the primary mission of the campus. Awards are made annually for outstanding teaching.

Enrollment at Edwardsville for the past eight years has annually averaged 10,000 students, with eighty-four Illinois counties, thirty states, and twenty-eight foreign nations represented. Nonetheless, SIUE is primarily a commuter campus, and most students live in communities within a 100-mile radius of the campus.

Background

Southern Illinois University at Edwardsville became interested in APEP in 1979 through the initiative of Earl Lazerson, then vice president and provost, and now president. The university administration and others knowledgeable about the project were excited about SIUE's selection as one of the ten pilot institutions by AASCU. The project afforded a unique opportunity for institutional self-examination.

During the 1978-79 academic year, a General Education Task Force was commissioned to begin the study of the general education program at SIUE. John Barker, professor of philosophy, was appointed chairperson of the task force. This group began its study of the general education program approximately six months before APEP activity began.

Since Earl Lazerson became president at about the time SIUE was notified of the APEP participation, Earl Beard, acting vice president and provost, became campus director for the project. Janet McReynolds, who had drafted the miniproposal, was selected as campus coordinator.

John Barker, Janet McReynolds, Earl Beard, and Earl Lazerson met to decide the appropriate structure for the project on the campus. It was decided that a separate committee should be formed to conduct the project. Subsequently, the project director and the campus coordinator met with the college deans to explain the project, to solicit their cooperation, and to request that they appoint a person from selected departments to constitute the committee. Initially, none of the committee members, including the campus coordinator, was given release time for the project. In retrospect, these two early decisions affected the committee both positively and negatively. The decision to keep the committee separate from other campus committees was probably sound.

The APEP committee membership on this campus was structured around departmental and school divisions and included faculty members from English, philosophy, business, speech communication, mathematics, biology, education, fine arts, sociology, and nursing. John Barker, chairperson of the General Education Task Force, agreed to serve as ex officio member of the APEP committee, as one way of eventually integrating the work of both committees. John Reiner, director of institutional research, agreed to serve on the committee to assist the committee in gathering necessary statistics. The committee also felt that a professional with a background in tests and measurement was vital to the project. To this end, the campus director attempted to recruit campus specialists who might assist the committee.

Implementation

Defining Purposes

To help direct its activities, the committee decided first to delineate AASCU purposes and SIUE purposes:

AASCU Purposes

- Assist ten institutions in defining conceptually and operationally four or five generic cognitive skills as

generalizable outcomes of the baccalaureate degree
- With the assistance of ten diverse state colleges and universities, conduct a pilot project on academic program evaluation with the intent of improving the educational content and the evaluation procedures for undergraduate programs
- Advance the state of the art in undergraduate program assessment.

SIUE Purposes
- Develop faculty and student awareness of the cognitive skills as generalizable outcomes of the baccalaureate degree
- Pilot test, and perhaps recommend an assessment plan for, systematic undergraduate assessment
- Encourage faculty members to examine courses in the curriculum with respect to the suggested definitions of cognitive skills, which will be helpful in

--discovering at what points in the university experience these skills are enhanced
--discovering appropriateness or inappropriateness of the committee definitions
--assisting curriculum committees to determine which skills are important to their curriculum, through both formal and informal means.

The members of the committee believed that the purposes, which should ultimately improve programmatic quality, could be achieved by
- improving awareness among members of the campus community about complex issues of systematic, programmatic assessment
- investigating currently available assessment instruments in order to ascertain appropriateness of local conceptual and operational definitions in comparison with commercial definitional assumptions
- serving as a sponsoring agency or forum to discuss issues related to undergraduate education.

The SIUE-APEP committee efforts were launched in January 1980 by the Washington project director. The stages of the model conceptualized at that time included a definition stage, a stage specifying acceptable levels of performance, a stage calling for the construction of an evaluation design, and an assessment stage. Participating institutions would independently determine the definitions and desired level of academic performance for each of the generic skills.

The project director's visits to SIUE generated several questions that subsequently appeared on other campuses involved in the project. Among them were the following: (1) Were SIUE and other participating institutions being asked to define generic skills needed or skills gained by undergraduates at the university? (2) What outcomes are expected of undergraduates? (3) Do all four-year programs enable students to demonstrate comparable generic skills? (4) To what extent should the APEP project be prescriptive, and to what extent should it be descriptive? (5) Is this project designed to be of a genuine experimental nature, and if so, why is there no control group? (6) To what extent, if any, will participating institutions be in contact with one another? (7) To what extent will the results of the testing be made public, and to whom will the data be made available? (8) To whom will the testing instruments belong? (9) What will be the result of this project, and how will it contribute to the literature of assessment; or will it duplicate the literature already available?

The committee planned carefully for a day-long retreat by drafting a statement of procedures and outcome expectations for the retreat. The expected outcomes for the day included: (1) the construction of a brief sketch of the definitions for each of the five skills, (2) an understanding of the various committee members' areas of agreement and disagreement about the generic skills, and (3) an identification of specific areas for additional reading and research.

In preparation for the retreat, the campus coordinator formulated a set of issues that the committee needed to confront:
- Will the SIUE committee include three. four, or five skills in the project? Should the possible existence of other skills be considered?
- How can faculty members not on the committee be involved meaningfully in the project? To what extent should there be general faculty involvement? How should the committee release and disseminate information to the faculty? What steps can and should be taken tc allay faculty fears about program evaluations?
- Should students be involved in decision making by the committee, and if so, how? How should students be involved in the assessment process?
- Can the project committee design an instrument that assesses the generic skills competence acquired in the major? What will the assessment scope be for the SIUE committee?
- Should the project committee inform the university of its findings? If so, by what means? Should the project committee attempt to influence the course of events at SIUE? If so, how?

According to the paradigm, the stages for the APEP project included the following: (1) faculty definition of desired intellectual abilities, (2) faculty definition of performance for each ability, (3) identi-

fication of where and how the abilities are taught, and (4) administration of multimeasure assessment instruments.

Formulation of the Definitions

The retreat proved profitable for the committee. The work of one committee member raised many questions about the cognitive skill of synthesis. Does creativity need to be defined? Does synthesis occur on different levels? Are analysis and synthesis really separable? Is there one skill that encompasses both synthesis and analysis? How would that one skill differ from evaluation? Can the generic skills be defined conceptually without considering the operational definition? To what extent do the attitude and educational philosophy on the instructor lead to the development of generic skills? An early decision made by the committee was to combine analysis and synthesis. The skill combining these processes was termed "critical thinking and reasoning." The Southern Illinois University at Edwardsville committee also decided that each of the generic skills definitions would be viewed as having two parts: the intellectual concept and the operational or behavioral definition.

Stage I was the development of conceptual and operational definitions for the generic skills. One of the generic skills, communication, generated vigorous debate. Some committee members contended that if communication were defined too broadly, its meaning would be weakened. Additionally, many questions about communication arose. For example, what kinds of organized systems can be considered communication? Does communication involve the intent of the speaker to ask an audience to make an inference or receive an experience? Does communication require that two persons or groups be consciously and overtly engaged in transmitting and receiving information?

Much of the work of creating two-part definitional statements for the skills was done through subcommittees that included members from several disciplines. This sharing of ideas with colleagues from other disciplines and with the university as well as colleges from other institutions provided opportunities for professional growth and development.

As the subcommittees developed their ideas, it became clear that SIUE definitions would be comprehensive. There was no attempt at this stage to determine those cognitive skills needed by students or the performance levels for generic skills competence. However, the Southern Iillinois University at Edwardsville definitions did attempt to include the widest possible range of skills.

Soliciting Comment and Support

After the local definitions had been formulated, revised, and carefully reviewed by the committee, the definitions were forwarded to all faculty members in the university, with instructions to return comments to the committee. Approximately ten faculty comments were received. Additionally, the definitions were published in a campuswide newsletter in an attempt to stimulate dialogue on the generic skills. A call for position papers was issued on the same topic; five were received.

Because several of the AASCU/APEP committee members were also on the contemporaneous General Education Task Force, and because the task force had created a subcommittee that also was defining the skills necessary for students in the general education program, there was considerable debate among task force members about necessary skills. The skill debated most was communication, with the debate reflecting the two viewpoints prevailing on campus: communication is a formalized presentation of opinions and values, and communication is the two-way development of interpersonal relationships. Some compromises were necessary; the general education definition of communication was more closely aligned to the presentational formal exchange, whereas the AASCU/APEP definition was more closely aligned to the interpersonal view.

In retrospect, it appears that if the impetus continues beyond the project's termination date, there should be a renewed effort to gain broader acceptance of the generic skill definitions. Additionally, there should be a reworking of the generic skill definitions in order to create an orderly hierarchy and a clear priority for assessing the skill subcomponents.

The SIUE AASCU/APEP committee spent one full quarter-year on the determination of the assessment mission, the selection of scope, and the design of the assessment plan. Each decision at this stage was related to several other decisions; consequently, the process was reexamined several times. Because this portion of the paradigm was emerging and the campus directors and coordinators were both learners and implementers, this stage was beset by the greatest uncertainty.

As mentioned previously, it was in Stage II that the need for a hierarchical structure for the skill definitions became apparent. Inasmuch as the SIUE definitions were comprehensive and global, it was clearly impossible to select all aspects of all skills for assessment. Consequently, it was necessary to establish priorities for assessment of the skills.

Each aspect of the decision-making process in Stage II was constrained by time. The decision to create

assessment instruments or to use ones already available was made (primarily) on the basis of the time constraints of the SIUE faculty--none had release time for the project--as well as on the basis of the project completion date. Additionally, as the scope of the project broadened from general education to the total undergraduate experience, the complexities of Stage II broadened, presumably because generic skills are learned throughout the entire undergraduate experience.

The SIUE committee decided on the following assessment and program portrayal plan:

- A random sample of 150 freshmen and 150 seniors was generated by computer. The freshmen sample pool consisted of entering students admitted by September 1, 1981. The senior sample pool consisted of students who had completed 144 quarter hours by July 1, 1981.
- Students from various general education courses were assessed. Course selection was based on enrollment, a two-hour time block for the class, and the cooperation of the instructor. Students were assessed in the following general education classes: Oral Communication of Ideas, College Algebra, Freshman Composition, Critical Thinking, Introduction to Art, Statistics, and Introduction to Sociology.
- All faculty members were mailed a questionnaire entitled "Pilot Project of the American Association of State Colleges and Universities." The questionnaire was accompanied by a cover letter from the vice president and provost explaining that the purpose for the questionnaire was to gather program information but not to "evaluate" or compare programs with one another.

Each student received a letter from the president of the university requesting participation in the study and explaining its nature and purpose. After this mailing, the project coordinator sent a form letter to the students announcing test dates and explaining how to sign up for the testing. Testing was conducted November 1 to December 5, 1981.

The assessment instruments consisted of multiple-choice tests in quantitative thinking and critical thinking and reasoning and one open-ended essay test based on a video presentation designed to assess the skills of communication and evaluation. The portrayal questionnaire was a multiple-choice instrument that asked students a variety of questions about class time, testing in their classes, classroom activities, assigned work, and instructor emphasis on the generic skills as defined by the committee.

All subjects were selected randomly and completed the assessment instruments and the portrayal questionnaire at one sitting prearranged not to interfere with their class schedules. Subjects in the general education classes completed the assessment instruments within the class but were given the portrayal questionnaire to complete and return the following day.

The SIUE committee decided on the following research hypothesis:

> Senior students will demonstrate significantly greater ability to use generic skills than freshman students.

The corresponding null hypothesis stated:

> There will be no significant difference in demonstrated ability to use generic skills between freshmen and seniors.

The program portrayal questions formulated by the SIUE committee were:

- What perceptions are held by selected faculty regarding the opportunity for students for students to learn five generic skills in the courses they teach?
- Does any particular cognate cluster of courses differ significantly from any other cluster regarding these perceptions?
- What perceptions are held by students enrolled in courses, regarding the opportunity for students to learn the five generic skills?
- Does any particular cognate cluster of courses differ significantly from any other cluster regrading these perceptions?
- To what extent is there agreement in perception (regarding the teaching and learning of the generic skills) between the faculty and students overall and within clusters?

The SIUE AASCU/APEP committee was unwilling to establish performance levels at the end of Stages I or II. In part, this resulted from the committee's view that the pilot project was a descriptive, trial experience. Additionally, the establishment of performance levels apart from specific instruments seemed to be arbitrary. The committee believed that desired performance levels could be recommended after the trial experience.

The committee had the greatest difficulty selecting scope, skill subcomponents, and the content domain for Stage II. The design of the assessment process at a large university is also a complex consideration. The committee considered several options before finalizing its plan. Among those options were:

- to sample core courses required in all or some of the academic areas having fifty or more majors in the program (thirty disciplines had fifty or more majors

in FY 79-80)

- to select a sample of twenty graduating senior's transcripts from the pool of thirty academic areas having fifty or more majors
- to choose two courses from seven of the thirty disciplines having fifty or more majors
- to administer the portrayal questionnaire to selected graduating seniors
- to sample the general education program by administering the portrayal questionnaire in fourteen general education courses having an annual enrollment of 500 or more.

The committee decided to undertake program portrayal with the belief that the task was necessary in order to have a clearer understanding of the meaning of the assessment results. Though the committee does not regret its decision to undertake program portrayal, this portion of the project proved slow and difficult. Program portrayal on the campus was obtained by two means--the college student experiences questionnaire developed by C. Robert Pace at the Laboratory for Research on Higher Education at UCLA, and an SIUE questionnaire designed to elicit attitudes and perceptions about teaching and learning generic skills.

The committee spent a considerable amount of time designing, revising, and pilot testing a combined faculty/student questionnaire for program portrayal. The program portrayal questions listed above guided the design of the portrayal instrument.

Selection of Measures to Assess Generic Skills

The SIUE committee decided in Stage I that analysis and synthesis were inseparable cognitive processes. As decribed above, critical thinking and reasoning emerged as the integrated skill description for analysis and synthesis. The committee examined alternative commercial instruments and considered devising instruments locally. However, the overwhelming magnitude of the task was such that the members were unprepared to design the instruments, and the deadlines were too short to do the task properly.

After many meetings and much deliberation, the committee decided to select analysis and synthesis questions from the ETS pilot tests with the highest reliability and test the items with several groups of SIUE students during the summer quarter of 1981. The pilot testing led to a few revisions. By fall 1981 the critical thinking and reasoning test consisted of eleven items, ten of which came directly from ETS.

After examining many tests for quantification, the subcommittee on quantification, consisting of faculty members from philosophy, business, and mathematics, decided to prepare an instrument probing these concepts: ratio, percentage, compound interest, geometric figures, equations (simple, linear), set inclusion and exclusion, statistics (mean and median), approximate numbers, distance-rate-time, and probability. Each member of the subcommittee prepared at least one problem for each of the ten concepts. Eventually, twenty-eight problems representing a wide range of difficulty were selected, and the time allotment was set at thirty minutes.

After the pilot testing, it was evident that thirty minutes was inadequate. Item analysis showed that some individual problems correlated negatively with overall achievement, presumably owing to faulty wording, a misleading diagram, or the like. Some questions were revised and others omitted, resulting in a test consisting of twenty questions.

A second pilot testing occurred with the new test in a senior mathematics class with seventeen students. The scores ranged from four to thirteen correct answers with a mean of 8.6 and a standard deviation of 2.6. An item analysis revealed a positive correlation between items and overall achievement on the test.

The quantification test was the first test administered during the testing sessions. No student answered all twenty items correctly; no one answered all twenty items incorrectly. The time allotment of thirty minutes was adequate.

Because the committee was opposed to a standardized test for communication, it set about to investigate alternatives. The subcommittee chose to assess communication and evaluation together by asking students to evaluate the effectiveness of messages in a videotaped debate. This seemed an improvement over more static and artifical writing and evaluation assessment instruments. Consequently, a video excerpt of fifteen minutes was extracted from a debate between David Rossen and Ralph Nader on nuclear energy.

As might be expected, the issue of how to evaluate the open-ended questions in two generic skills--communication and evaluation--proved to be an extremely complex problem. Students were asked five open-ended questions after viewing the video presentation. They were given the opportunity to take notes. The pilot testing of the communication/evaluation instrument took place during the summer 1980, and several questions were rewritten as a result.

Data Collection

The assessment design chosen by the committee

involved the random selection of 150 reshmen and 150 seniors. As with other institutions that had used the random sampling procedure, considerable attention was given to the proper means of soliciting student cooperation. A gift package comprising campus tickets, passes, and coupons approximately equivalent to ten dollars was offered to students for participation. In addition to receiving a letter from the president of the university and the campus coordinator, students were telephoned by faculty members on the committee. In spite of all efforts, only forty-two freshmen and twenty-nine seniors actually participated from the randomly selected pool of 300 students.

The committee members had modest success in obtaining the cooperation of faculty members who teach general education courses. As a result, 267 general education students completed the testing during the class sessions. Thus, the total SIUE sample was 338 students. The faculty program portrayal instrument was sent to 536 full-time faculty members in the fall of 1981. One hundred seventy were returned.

Data Analysis

The communication/evaluation results were analyzed according to a holistic scoring procedure adapted from the College Board testing program entitled *The English Composition Test with Essay*. Two persons unrelated to the project were trained by the committee to score the student essays according to agreed criteria. Two scores were generated: a communication score and an evaluation score. The critical thinking and reasoning and the quantification tests were analyzed separately in order to answer the questions formulated by the committee. The college student experiences questionnaire was analyzed according to the subject categories in the instrument. All data were placed on computer tapes to facilitate statistical analysis.

Outcomes

Generic Skills Assessment

The SIUE/APEP committee recognized from the beginning that the most appropriate research design to probe generic skill improvement of college students would be a longitudinal study. Because difficult time and budget constraints precluded such a choice. however, compromises had to be made. The committee's choice of studying concurrent groups of randomly chosen freshmen and seniors along with certain general studies classes still provided significant, if less than surprising, clues to the effects of the college experience on SIUE students. Comparisons were matched to control for ACT scores and age. Results of interscale correlations generally supported the validity of individual scales, with the one caution that "communication" and "evaluation" scores--derived from a written instrument--may be measuring the same skill rather than distinct skills. Reliability of the scales was shown to be high.

In each skill category (critical thinking, quantification, communication, and evaluation) scores were significantly higher for the random senior group than for the random freshman group, with the largest gain shown in quantification. General studies freshmen scored somewhat higher than random freshmen on all tests except critical thinking. After controlling for age and sex, year in school continues to be positively related to test scores, but slightly so. The difference between freshman and senior level skills--although small--persisted after controlling for ACT scores. The data supports the hypothesis that students can be expected to experience moderate but consistent improvement in generic skill level while at SIUE, although analysis of the data in light of ACT scores discloses that college attendance does not result in a major equalization of skill levels.

College Student Experience Questionnaire

The data generated by the Pace instrument seemed to confirm that the student sample was demographically representative of the SIUE population and provided a close "fit" with the university's existing self-image. In short, the results were consistent with what might be expected of an urban commuter school; students reported relatively low involvement with the university other than class-related activity. However, on about half of the activity/involvement items--including library use, faculty content, course learning, clubs and organizations, personal development, and discussion of topics in conversation--seniors reported significantly higher levels of participation than freshmen. In a broad array of areas, seniors are somewhat more involved with the university than are freshmen. A revealing section of the instrument asks students for a self-report on gains in eighteen skill and appreciation categories. Students were highly positive in eight of the eighteen areas and perceived moderate gains in four others. In the context of AASCU's APEP, it is significant that some of the larger self-reported gains occurred in areas relating to communication (writing), synthesis (putting ideas together), and evaluation (understanding self and others and developing values and ethical standards).

Moderate gains were recorded in analysis (thinking critically and logically) and quantitative thinking. In virtually all academic skill areas, seniors exhibited substantially greater gains than did freshmen.

Course/Program Portrayal

The committee designed its own instrument to portray faculty and student estimates about where, when, and how well generic skills are addressed in the curriculum. The faculty, not surprisingly, considered SIUE's courses more helpful than did students in all skill areas except those related to quantification. Faculty members thought their courses were most helpful in developing skills in critical thinking and nearly as helpful in communication and evaluation subskills. They were least positive about the usefulness of their courses in quantitative learning. Students, on the other hand, ranked quantification skills as most helpful, but differences among skill areas were small for students. Student and faculty perceptions diverged on the helpfulness issue nearly across the board. Also, faculty generally saw their courses as emphasizing critical thinking more than fact recall, whereas students viewed the courses the opposite way. Finally, the data reflecting emphasis on skills were analyzed in light of faculty disciplinary groupings. Variation among emphasis patterns was found to be so large that the committee was forced to conclude that a sound general education or distribution-among-disciplines program is necessary if students are to be uniformly exposed to those skills in their educational programs at SIUE.

Primary Outcomes of the Project

1. We now know more about the skills profile of SIUE students. The college experience does seem to make a difference, and the institution appears to be fulfilling its mission to the community and, most important, its responsibility to the students who place their trust in the university experience. Still, some qualifications and reservations are necessary. We don't know precisely, for example, how much change is, attributable to normal maturation processess, how much to different abilities of current freshmen compared with freshmen of three to four years ago, and how much to the simple fact that grading and academic suspension practices of the university "select out" less skilled students, presumably leaving a more skilled senior class from which to select subjects.
2. We now know how valuable it is for faculty members from diverse disciplines to engage in a continuing, basic dialogue about what a college education means. Much controversy and more than a little impatience were evident as we discovered we often differed widely on what students should know and be able to do. Perhaps more frustrating, we found ourselves in frequent disagreement about pedagogical values, styles, and philosophies. Faculty members perhaps assume too much about campus consistency on such issues when not faced directly with empirical evidence of diversity. In this group, frustrations and conflicts were usally translated into constructive decisions.
3. We now know how difficult it is to get faculty members to sustain commitment to a long-term project. This is especially true when the service is voluntary, involving frequent meetings, long hours, constant decisions, impending deadlines, and written reports.
4. We now better realize the importance of a consistent master paradigm for a project of this magnitude. On several occasions, the local APEP committee was unsure whether the original paradigm was still in force or a newer version was being implemented. For example, at one point the description of project intent seemed to change from a descriptive pilot study ("What does basic skill learning look like on our campus?") to a diagnostic/evaluative study ("How can these results be applied to suggest program modifications in the institution?")
5. We now know from program portrayal instruments much more than we did about the perceptions of faculty members and students concerning how and where the generic skills are addressed in classes. Further analysis of the data should yield interesting comparative conclusions; for instance, do students and faculty members in the same course (or program, or major) have similar perceptions of which skills are stressed? Do faculty members believe evaluation skills are learned in a class when students have not noticed their being emphasized or taught?
6. We now know much more about the development of "home-grown" instruments and their integration with standardized instruments into a coherent research design. Still, we must recognize how difficult this is. The desire to be creative and original led us to attempt to measure communication and evaluation skills by a blended-purpose instrument administered in tandem with a videotaped presentation of a recent debate on a sociopolitical issue. Trade-offs were obvious: the creativity and specificity gained were perhaps purchased at the expense of administrative

complexity and somewhat ambiguous validity. Yet many standardized instruments seemed impersonal, unsuited to the particular skill definitions generated in earlier phases of the study, and inappropriate to the SIUE student population.

7. We now can assume a somewhat heightened awareness among the SIUE community of the importance of generic skills assessment. The AASCU-sponsored project attempted in several ways to publicize its mission and to generate campus dialogue on this issue. Position papers were requested, committee members solicited aid and comments from colleagues, newspaper articles described project goals, and skills definitions were published in *The Observer*, a campus newsletter. We noted that even those persistent efforts were insufficient to stimulate significant response.

Conclusions

1. The APEP concept should become an ongoing concern at SIUE. The current committee believes that a university has a responsibility to be a continually self-monitoring, self-aware institution. Generic skills definition and assessment enacts that responsibility and assures our community of our integrity and willingness to seek feedback on effectiveness.

2. The university should appoint a committee to evaluate the short-term conceptual and research procedures of the APEP committee. Many decisions were made with difficult time constraints, limited budget, and ambiguous information about national project expectations. A new committee should bring an objective viewpoint to bear on these decisions. Some important questions: Which policies are worthwhile and should be maintained? Which methodological procedures were suspect? How might the decision-making process be streamlined? Should an APEP committee engage in more interpersonal "team-building" activities prior to its procedural phases? How could the assessment procedures be made more replicable?

3. The university should create a local Academic Skills Evaluation Committee as an ongoing subgroup of the faculty senate's curriculum council. This committee would supervise the investigation of generic skill learning at SIUE and initiate institutional policies relevant to that learning. It might start with the current APEP models and go beyond them. Careful balance of academic specializations and research preferences should be maintained on this committee, along with consistent student representation. Appointees should have a clear realization of the work involved and state a clear intention to follow through; the existing APEP committee did not encourage such a commitment and thus experienced absences, vacillating contributions, and even resignations from some members. The projected committee should become an on-load, assigned time for faculty members, and student members should receive some form of recognition for their participation.

4. The university should mandate through the Academic Skills Evaluation Committee a systematic, longitudinal study of generic skills learning. That is, a reasonably large but manageable group of freshmen (perhaps 200) should be identified and tracked throughout their college careers. Periodic assessment of their basic skill competencies might be accomplished through strategies similar to those of APEP. Additionally, such a longitudinal study should include

- data on informal college experiences similar to that generated by the Pace instrument. We should stay attuned to the extracurricular and cocurricular activities chosen by students, for these also affect the development of skills. Some studies indicate the informal network of college learning is ultimately more potent than classwork in the acquisition of life skills.
- a face-to-face interview component with appropriate self-report instruments. We know remarkably little about how skills learning in college feels from the student point of view. Narrative reports, journals, or interview responses might be content-analyzed to yield rich insights to supplement the hard data with a more human perspective.

5. The university should guarantee financial support and ongoing administrative commitment for the longitudinal study. To begin such a project without a clear commitment to carry it through could be worse than not starting the project at all--and would be a disservice to project participants and the university community. Although the present pilot project was more descriptive than diagnostic (we were not intending to use our findings to initiate curricular revision or recommend different educational philosophies), it is not unreasonable to expect that a four-year longitudinal study, thoroughly reasoned and coherently conducted, might result in such diagnostic recommendations.

6. The university should make an ongoing, systematic commitment to create campus dialogue on skill definitions and criteria for competence. A broader acceptance of and agreement on an orderly, hierarchical generic skill scheme would be helpful, and this might be accomplished through university

sponsorship of colloquia, seminars, informal publications, and open meetings. Several questions could be meaningfully addressed: What is the central purpose of the teaching-learning process? Have Southern Illinois University at Edwardsville students mastered the generic skills? Which behavioral and attitudinal changes are reasonable to expect of the Southern Illinois University at Edwardsville students? How can Southern Illinois University at Edwardsville facilitate generic skill learning to make it both efficient and humane?

Written By

Robert Anderson, John Farley, and Janet McReynolds

Institutional Contacts

Robert Anderson, professor of speech communication
John Farley, professor of sociology
Martha Welch, professor of nursing
Janet McReynolds, assistant to the vice president and provost

State University College at Potsdam (NY)

Institutional Profile

Potsdam College was founded in 1816 as a small academy that later became a Normal School in 1867. In the 1940s, the college became a Teachers College and in 1962, a college of Arts and Science in the massive State University of New York system. It is currently a comprehensive college with its strongest programs in music, computer science, mathematics, economics and education. The undergraduate population of the college is now 3,600 students with an additional 600 part-time graduate students.

Background

At the time that APEP was being proposed to member colleges of AASCU, administrators at Potsdam College were already concerned about intellectual skills. The college's most recent campus mission statement contained statements on the importance of student skills, and the staff had begun rewriting the mission statement and wondering whether these objectives really were fulfilled. A desire to understand better the outcomes of our educational programs through evaluation led to Potsdam's application to the AASCU Resource Center for selection as one of the ten state-supported colleges to participate in the APEP. We were particularly interested because the vice president for academic affairs and the associate vice president had both participated in the summer institute of 1978, when work began on skills development.

Implementation

In the fall of 1979, an already existing Committee on Academic Skills was abolished and a new Committee on Intellectual Skills was formed. It included key campus leaders and representatives from every significant academic area. This committee began work in October 1979 and quickly became a part of campus life.

Defining Generic Skills

The committee spent the first 2-1/2 months discussing and reaching consensus on a conceptual definition of the four skill areas to be included in APEP. These were analysis, synthesis, quantification, and communication. Faculty and administrative staff members alike found this experience to be highly stimulating, because they normally do not set aside time to sit and discuss educational outcomes from the perspective of a wide variety of disciplines. The committee did conduct these discussions and, in the process, determined that valuing was an additional skill that should be studied in the project.

In February 1980, the committee began reexamining the broad conceptual definitions with an eye toward defining subskills for each of these. The committee was subdivided across the areas of (a) natural science, (b) social sciences, and (c) humanities to discuss the subskills needed under each working definition. Although reaching a consensus occasionally required difficult concessions, this second task was achieved by the end of that spring. More importantly, this fusion demonstrated to the now-reunited committee how similar the separate concerns had been and how comparable they could become when formulating unified subskills.

Moving from Conceptualization to Operationalization

Early in the summer of 1980, a smaller but very representative version of the Committee on Intellectual Skills renewed its activities. This committee began the process of converting the definitions that had been previously determined into operational statements, choosing and modifying a research design, selecting appropriate evaluation formats, and choosing appropriate but varied evaluation instruments. This process of creating operational statements from the conceptual definitions was a difficult and frustrating task. Having spent months developing statements that characterized a valid institution-specific version of the generic skills, the members felt some ambivalence about writing simplified, testable examples of their formerly comprehensive works. This stage ultimately required careful editing of the lists that had been prepared, sometimes with disagreements among committee members but with a final consensus. The committee was not completely satisfied with its observables, but it understood as it entered the actual testing cycle that without the crucial

transition from conceptualization to operationalization, the project would not later measure change in predictable or meaningful terms.

Choosing a Research Design

In creating the research design, the committee adopted a model it labeled a "Multi-Cell Research Design." It was determined by the group that the best way to measure skills might be across class years and across major academic divisions of the college. These subdivisions were liberal studies natural science, liberal studies social science, liberal studies humanities, Crane School of Music, Educational Certification Program, and undeclared students. We had been discussing the possibilities of purchasing tests, when the committee decided to call the APEP consultant, Jonathan Warren from the Educational Testing Service. Warren visited us in the summer of 1980 and explained to us the instruments he was developing as well as the choices open to us for our own observations. One choice included working with him to aid his test development, using existing instruments. Warren had been working with colleges in California but said he would welcome APEP colleges and could ultimately provide us with data about our students.

Developing and Administering the Tests

We later determined that a combination of Dr. Warren's four tests and some locally developed instruments would be the best way to go. For example, two art historians developed a computer-scorable exercise for analysis/synthesis. A history professor produced a one-act play, *The Birdbath*, after which students wrote an essay concerning receptivity and sensitivity, an exercise in both communication and valuing. During the fall of 1980, we began to draw up the lists of students who could potentially participate in the testing exercise. We also developed questionnaires to collect data about these students concerning their cocurricular learning, recognizing that students can easily perfect some of these skills in cocurricular activities. In addition, we asked students which of the courses they had taken had taught them the skills we were examining. We asked faculty members for these students to rate them on the five skills and to explain how they develop courses which lead to improving these skills. In the spring of 1981, with only the fact that they were helping the college as their incentive to appear, students were assembled and spent an entire day in the testing exercises. A total of 177 students actually completed all exercises.

Outcomes

During the summer, we began the process of tabulating information, and a group of readers spent an entire day away from the campus at a retreat center reading the essays and scoring them on the basis of communication as well as valuing. By the end of that second summer, the committee then had information ready to be studied. We were then able to make comparisons among the various cells of our research design. The four tests from ETS were scored by ETS, and Warren returned to the campus to discuss results with us. The fact that Potsdam students had performed well in his tests helped strengthen our knowledge that this had, in fact, been a worthwhile undertaking. Each of Warren's visits also rekindled enthusiasm in the Potsdam committee members.

The steering committee then became interested in further development of the locally produced test instruments. The study showed us in which year and in which academic area students scored better on these skill areas.

Conclusion

We have only slightly tapped the wealth of information we collected. We need to find time and resources to continue to analyze the data and to raise significant questions about our curriculum.

Written By

A. Paul Loucks and Bruce A. Conroe

Institutional Contact

Bruce A. Conroe, vice provost
P. William Shearer, associate professor of psychology

University of Nebraska at Omaha

Institutional Profile

The University of Nebraska at Omaha (UNO) is one of three major administrative units of the University of Nebraska. The institution was founded as a private university in 1908 and became the Municipal University of Omaha in 1931 and part of the University of Nebraska in 1968.

Organizationally, the University of Nebraska at Omaha includes six collegiate units, which report to the vice chancellor for academic affairs. These colleges are Arts and Sciences, Business Administration, Continuing Studies, Education, Fine Arts, and Public Affairs and Community Service. In addition, the University of Nebraska-Lincoln's College of Engineering and Home Economics offers programs on the University of Nebraska at Omaha campus. Graduate programs are offered through the system-wide Graduate College of the University of Nebraska.

Currently, nearly eighty baccalaureate degree programs (including those in engineering and home economics) are available at UNO. Graduate degrees available on the UNO campus include the Master of Arts, Master of Science, Master of Business Administration, Master of Public Administration, Master of Social Work, Master of Professional Accounting, Master of Music, and Specialist in Education. In addition, the Department of Psychology participates in a joint Ph.D. program with the University of Nebraska-Lincoln.

The University of Nebraska at Omaha is the only comprehensive public university in a metropolitan area of approximately 600,000 people. A commuter institution having no dormitories, UNO draws virtually all its students from the Omaha metropolitan area. The student body of approximately 15,000 is almost equally divided between men and women. Minority students account for an estimated 10 percent of the student body. Approximately 75 percent of UNO students are employed on either a full- or part-time basis, and the median student's age is twenty-six. The average composite ACT scores for 1981-82 entering freshmen was 19.0.

UNO has a full-time faculty of nearly 400. Over 75 percent of the faculty members possess terminal degrees, and many are actively engaged in research and public service activities as well as instruction.

Background

Participation by UNO in APEP was presaged by developments in three key areas of campuswide significance: general education, planning, and program evaluation. First, during the 1978-79 academic year, each college reviewed its general education requirements, a process undertaken upon the initiative of University of Nebraska President Ronald W. Roskens. It should be noted that there is no campuswide, or systemwide, general education requirement at the University of Nebraska; rather, each college within the university establishes its own requirements. At UNO, the 1978-79 review largely reaffirmed existing requirements; however, it did generate widespread interest in the general education component of the curriculum and served as the catalyst for an extended study of general education requirements by the college of arts and sciences.[1]

Simultaneously, considerable faculty and administrative attention was devoted to planning. Early in 1979, the development of college-, campus-, and university-wide five-year plans was initiated. The five-year plans were to be prepared in accordance with a set of twelve planning guidelines adopted by the board of regents. A format for the UNO college and campus plans developed under the aegis of the Ad Hoc Academic Program Steering Committee included an evaluative component. The nascent Academic Program Evaluation Project was viewed as a promising approach to satisfying this evaluation requirement, and reference was made in the draft campus plan to the possibility of UNO's participation in the proposed, ten-member APEP consortium. This planning effort eventually was superseded by a more centralized process; nevertheless, interest had been stimulated in APEP as an evaluative tool.

Finally, the newly appointed vice chancellor for academic affairs, Otto Bauer, initiated discussions in July 1979 about the formation of an academic planning council, to be charged with conducting an ongoing qualitative and quantitative review of all of UNO's academic programs. Whereas this review process was conceived as focusing on departmental units and their majors, APEP was seen as offering another, complementary dimension for analyzing the undergraduate curriculum by concentrating on generic intellectual skills.

In brief, APEP emerged at the University of Nebraska at Omaha when UNO was addressing general education, planning, and program evaluation. The timing of the project, as well as its substance, coincided agreeably with campus initiatives in these important areas.

Implementation

In launching APEP at UNO, the initial task was to select the faculty committee that would guide the project to completion. Essentially two alternatives were considered--appointing a new faculty committee to assume responsibility for the project or adding APEP to the charge of an existing body. After a review of the alternatives and their implications, it was decided to invite the Educational Policy Advisory Committee (EPAC) to serve as the APEP faculty committee.

At the time APEP was initiated, EPAC had served as UNO's campuswide curriculum committee for nearly six years. The committee consisted of one faculty member from each of the colleges of arts and sciences, business administration, education, fine arts, and public affairs and community service and from the library; one undergraduate student; and one graduate student. Faculty members served overlapping terms of three years, whereas student members served one-year, renewable terms. Two faculty members rotated off the committee each year. The committee reviewed new courses and programs and advised the vice chancellor for academic affairs about a wide range of academic policy matters. Meetings were held twice a month and chaired by the vice chancellor.

Although a separate committee would have been able to focus exclusively on APEP, some concerns had been expressed about the possible proliferation of committees. Given its involvement in curricular matters, its representative composition, and its internal rapport, EPAC was the logical choice. After reviewing APEP materials, the members unanimously agreed to undertake the project.

In 1979-80, faculty members of EPAC included Dr. William Callahan, Associate Professor, Special Education; Dr. John Kasher, Professor of Physics; Dr. Janet Rives, Associate Professor of Economics; Mr. Tom Tollman, Assistant Professor and Reference Librarian; Dr. Sam Walker, Associate Professor of Criminal Justice; and Mr. Bob Welk, Assistant Professor of Dramatic Arts. The following year Professors Callahan and Welk completed their terms on the committee. They were replaced by Dr. Robert O'Reilly, Professor of Educational Administration; and Dr. Warren Prince, Professor of Music. In 1981-82, Professors Kasher and Rives were replaced by Dr. Bruce Baker, Professor of English, and Dr. Richard Ortman, Associate Professor of Accounting. Dr. Deanna Finkler, Associate Professor of Psychology, served as consultant.

The committee first addressed the scope of the project (i.e., whether the focus should be on general education or on the general outcomes of a college education). The prevailing view was that the five APEP skills, although closely associated with general education, were by no means limited to general education. Had the project been confined to general education, a complicating factor would have been the lack of uniformity among the colleges in their general education requirements. This was not, however, the reason for extending the scope of the project beyond general education. Rather, it was believed that the skills were developed at all levels of the curriculum and that this approach would afford an analysis of the effectiveness of the entire curriculum in the development of the five key intellectual skills.

The broad goals of the project also were identified at an early date. Indeed, agreement on project goals proceeded hand in hand with the selection of the scope of the project. The two basic goals established for the project were (1) to discover the level of proficiency in the five APEP intellectual skills displayed by students nearing completion of a baccalaureate degree and (2) to determine whether graduating students in the respective colleges demonstrated different profiles of the generic skills. The profiles were to be made available to colleges for their use in considering possible curricular changes. It was recognized, however, that primary responsibility for initiating curricular change rested with the faculties of the respective colleges. Moreover, given the differing curricular emphases among the various colleges, as well as the absence of a campuswide general education component, some differences in the profiles among graduates of the several colleges were expected. Finally, it was emphasized to the faculty that APEP was not intended to serve as a review of individual programs nor as a tool for reallocating resources within the institution. Although this evaluation approach could be adapted to those purposes, University of Nebraska at Omaha's goals were more modest. Essentially, it was interested in developing a diagnostic approach; the prescriptive remedies, to the extent they might prove necessary, were to be outside the scope of APEP as it was conceived at UNO.

Defining the Generic Skills

Having charted the broad goals of the project, EPAC turned to developing campus-specific definitions of the generic skills. Discussions centered first on communication, quantification, analysis, and synthesis; the committee agreed to address valuing after completing its work on the other four skills.

To stimulate discussions, several handouts were distributed to committee members. Documents included the April 1979 issue of *Alternatives*, which contained several articles on intellectual skills; draft definitions of intellectual abilities, excerpted from a case study used in Fall 1979; regional seminars conducted by the Resource Center for Planned Change; and a summary of key elements of Bloom's taxonomy of educational objectives.

Finkler also prepared matrices relating skills, content areas, and modalities to Bloom's academic skill levels. Subsequently, additional materials on Bloom's taxonomy were provided, along with a description of Kohlberg's model of the development of moral judgment and a copy of a sample student's responses on the Rest Defining Issues Test.

After considering several ways of defining the skills (e.g., global statement or an exhaustive list of all components of each skill), it was decided to adopt Jonathan Warren's definitions of communication, synthesizing, and analytical skills as a point of departure. These definitions had been published in the April 1979 issue of *Alternatives*. Warren's approach seemed particularly well-suited to the project, in that instruments being developed by the Educational Testing Service would be based on those definitions. Moreover, since they were stated in terms of three qualitatively distinct skills (low, medium, and high), Warren's definitions effectively combined two steps of the APEP paradign (writing definitions and establishing skill levels) into one. Most important, Warren's definitions were viewed as clear, thorough, illuminating, and well-reasoned. Some consideration was given to abbreviating these definitions; however, the committee decided to adopt the published versions.

Attention then turned to developing a definition of quantification following the three-level approach of Warren's definitions of communication, synthesizing, and analytical skills. Abstract, curriculum-based, and tool-based draft definitions of quantification were developed. Discussions centered on the abstract and curriculum-based definitions; elements of the two were synthesized into the definition adopted by the committee.

Valuing was the last skill to be considered. Dr. William Blizek, Associate Professor of Philosophy and a recognized campus authority on valuing, provided helpful insights during the initial discussion of this skill. A number of key ideas relevant to values and valuing were considered. These included intellectual courage, intellectual openness, intellectual tolerance, sensitivity to other persons and their sensitivities, social responsibility, intellectual honesty, searching for truth, and critical (i.e., discriminating) thinking about values. Although committee members approached valuing from different perspectives, they concurred that it is a vital outcome of the undergraduate curriculum and should be addressed in a project such as APEP. A three-level definition centering on the concept of critical thinking about value-laden issues was adopted.

Upon reflection, it was agreed that a succinct definition should be developed for each of the five skills in addition to the previously adopted statements describing the three levels of performance. This brief task, which lasted approximately four months, concluded the definition-writing stage of the project. In addition, members of the committee arranged for pilot-testing in their classes of measures being developed in conjunction with this project by the Educational Testing Service.

Designing Testing Plans

During the summer, a tentative plan for the testing phase of the project was developed in some detail. This was accomplished by the project consultant, who worked the equivalent of one month for this purpose. To obtain additional guidance, Finkler arranged to meet with a faculty member of each of the colleges whose students were to take the APEP test. The faculty representatives, designated by the deans, typically were members of their college's curriculum committees. This approach was intended to identify specific college purposes for the project, in addition to the more general ones adopted by the committee, and to elicit problems peculiar to the students of individual colleges. These contacts also provided some helpful feedback regarding faculty concerns and perceptions of the project.

Two meetings were held with each college representative. During the first, Finkler described the project in some detail and posed a series of questions for further discussion. During the second session, these questions were considered and the representative's counsel sought. Among the issues discussed were the adequacy of the generic skill definitions from the college's perspective, the preferred measurement strategy, the purposes the study might serve for the college, and the college's

preference in the relative emphasis on program ... aval and student outcome assessment.

The tentative plan was organized around three of the five charts made available at the June APEP meeting. The charts proved particularly helpful in articulating relationships between project purposes, the relevant student populations, and student sampling procedures. In addition, the project consultant prepared extensive notes on the then-current ETS measures of quantification, communication, analysis, and synthesis. The tentative plan was available at the beginning of the new academic year for review and discussion by EPAC.

Selecting Testing Instruments

During the fall semester of the 1980-81 academic year, EPAC devoted considerable attention to reviewing testing instruments (particularly the ETS measures) and considering sampling procedures and other issues raised in the tentative plan. During the committee's review of measurement instruments, the results of the pilot tests of the ETS measures became available, together with a set of revised ETS scales that reflected the results of the recently completed pilot testing.

Working individually or in pairs, committee members examined the ETS scales in depth. After extensive deliberation, the committee agreed on the following measures:

Measure	*ETS Form*	*Items*
Analysis	31	6, 9
	32	3, 6, 8
	33	1, 5, 7, 9
Quantification	27	1-9, 11-14
	21	22
	22	17
	23	20
	24	7, 20
Synthesis	44	1-12 (all)
	41	10
Communication	14	1-12 (all)

These four measures, consisting of fifty-two multiple-choice items, constituted the first part of the UNO APEP instrument.

The committee determined that, to accomplish its purposes, the test for communication skills should be supplemented by a writing exercise. After considering several possibilities, the committee decided to adopt an essay test developed in consultation with the Department of English and based on its diagnostic writing examination. This approach was particularly appealing because the department has, in recent years, developed a highly refined method of scoring essays for the purpose of placement into freshman composition courses as well as determining whether certain students meet the writing proficiency requirements for graduation. This alternative had the additional advantage of being linked to an established curricular requirement. The essay test consisted of five questions on relatively current social or political issues and were considered to represent a medium level of difficulty. Students were to select one of the five questions.

For valuing, the committee decided to adopt the Defining Issues Test (DIT) developed by James Rest. It was generally agreed that this was the best alternative readily obtainable; the availability of regional norms added to the appeal of the DIT.

Designing Sampling Procedures

The committee next addressed the critical issue of sampling procedures. A key question was whether the test should be administered during class time or given outside class during specially arranged testing periods. The obvious advantage of classroom testing was that the desired sample size could be ensured; there would be a more or less captive audience, assuming the cooperation of the respective faculty members and students. The main disadvantage was that nearly a full week of class time would be required to administer the test. Consultations with a variety of faculty members suggested this alternative was not feasible.

Administering the test outside class posed a different set of considerations. By inviting specific students, as contrasted with classes, a more exact random sample could be developed. However, the rate of response to an invitation to take a two-hour test typically is not high. Hence, a specific number of examinees could not be ensured, and an excessively low turnout could impair the project seriously. Nevertheless, it was agreed to adopt this approach, despite its inherent difficulties, with the understanding that special efforts would be made to maximize the rate of response.

The committee agreed to a sample size of 100: fifty senior students (defined as those having ninety semester hours or more) from each college, plus fifty freshmen for comparative purposes. These were to be stratified random samples, drawn to reflect differences in age, sex, race, grade-point average,

and major. The office of institutional research assisted in drawing samples.

Administering the Test

Several steps were taken to ensure a response rate that would contribute to the success of the project. First, arrangements were made to administer the test at four different times during the week of April 6: Wednesday evening, Thursday evening, Saturday morning, and Saturday afternoon. It was hoped that interested students could fit one of these times into their schedules. To encourage student participation, it was decided to charge a fee of $12.50 for taking the test. Clearly, this proved to be a positive inducement, based on informal student reactions. Funding for this purpose was provided by the FIPSE-AASCU grant.

Invitations were mailed to the 300 students identified in the stratified random sample (fifty from five colleges, plus fifty freshmen). To assist in standard-setting, the deans were asked to identify students they would expect to perform at high and medium levels in each of the five skills. Approximately forty students were identified and invited to take the test, in addition to those randomly selected.

The response rate to the first mailing was nearly 40 percent, with about 80 percent of the replies being favorable. A second letter was sent to the students who had not responded to the first invitation within one week. Both letters were written in an informal style. Supplementing the mailings, EPAC members telephoned students who had not responded within eighteen days of the first letter. To provide positive reinforcement, a confirming letter was mailed to students who had indicated they would take the test. Ultimately 160 of the 340 invitees took the test, a participation rate of 47 percent. Approximately 10 percent of the students who responded favorably to an invitation failed to appear.

The test was administered in a large lecture hall in a familiar classroom-office building by a member of the project staff, who provided a general explanation of the project and answered questions at the beginning of the exam period. The staff member also read aloud the instructions appearing at the beginning of each test booklet. The four ETS measures were given first. Students were allocated fifteen minutes to complete each part of the test. After a ten-minute break, students were allocated one hour to complete the DIT and essay segments of the exam.

One of the examinees was a reporter for the student newspaper who, shortly after taking the test, wrote a major feature story about the test and the project. Occupying one-fourth of the front page of the April 15 edition, the article provided excellent campuswide publicity of APEP.

Setting Performance Standards

The establishment of performance standards on the APEP test was accomplished in two ways. First, EPAC members undertook individually to determine appropriate ranges for high, medium, and low scores on each of the ETS tests. Subsequent committee discussion revealed consensus among the members about the scores for each level of performance. Accordingly, the committee readily reached agreement on the following score ranges for the ETS tests:

Range (No. of Correct Responses)

Skill	No. Items	High	Middle	Low
Analysis	9	7-9	5-6	0-4
Quantification	18	14-18	9-13	0-8
Synthesis	13	10-13	6-9	0-5
Communication	12	9-12	6-8	0-5

Existing English department standards (A=90-100; B=80-89; C=70-79; Not Passing=0-69) for the essay diagnostic test were adjusted to compensate for what was considered insufficient time for this portion of the test. The standards ultimately employed are as follows: A=85-100; B=75-84; C=65-74; D=50-64; not passing=0-49. No standards were established for the DIT.

Next, each college was asked to indicate its expectations for performance on the test by its senior students. Typically, this task was accomplished by the college's educational policy committee. Each college indicated the percent of its senior students expected to score in the high, middle, and low ranges for each of the ETS tests, and for "A," "B," "C," and "Not Passing" scores in the essay test. While colleges were to provide copies of the tests, they were asked to state their expectations in the absence of information about their students' performances on the measures in order to avoid biasing their judgments. One college's expectations are shown below as an example of how this task was approached:

Anticipated Performance (Percent)

Skill	High	Middle	Low	
Analysis	48.3	46.7	5.0	
Quantification	35.0	56.0	9.0	
Synthesis	45.0	48.0	7.0	
Communication	56.0	41.0	3.0	
	A	B	C	Not Passing
Essay	17.0	45.0	38.0	0

Analyzing the Data

The test data were analyzed during the summer by the project consultant, who prepared an extensive technical report on the measurement procedures and results. Some of the more salient findings of the report are summarized here.

Two methods of analyzing the student test scores were used. The first method compared proportions of students achieving scores in the high, middle, and low proficiency ranges identified by EPAC to distributional profiles established by the collegiate educational policy committees as representing both "desirable" and "reasonable" levels of achievement for their senior students. Chi-squared tests were used in this analysis.

The analysis revealed that, in general, the student test scores differed from the expectations of the faculty committees. Scores on the four ETS measures were both lower and higher than faculty expectations, depending on the college. In three of the five colleges, for example, scores on the quantification test were higher than expected. In two colleges, students scored higher than the faculty expected on at least three of the four ETS measures. However, it should be noted that these results were as much a function of faculty expectations--which vary across the colleges--as of student performance. In contrast to the ETS measures, student scores on the essay test were consistently lower than faculty expectations. By this method of analysis, seniors scored higher than freshmen on the quantification, synthesis, and essay tests.

The second method of data analysis addressed differences: (1) among seniors in the five colleges and (2) between seniors as a group and freshmen. For each comparison, a multivariate analysis of variance was performed. The results of these analyses indicated that only the essay and DIT "P" scores differed significantly among seniors in different colleges, and that only the essay scores differed between seniors and freshmen.

The analyses showed that seniors in the colleges of arts and sciences and business administration scored higher than those of the other colleges on the essay, and that arts and sciences seniors, scored highest among the colleges on the DIT. Seniors scored significantly higher than freshmen on the essay; however, neither group's mean score represented a passing grade of "A," "B," or "C." Although the means on the ETS measures did not discriminate among seniors in different colleges or between seniors and freshmen in this method of analysis, the means were predominantly the middle to high ranges as identified by EPAC. Although standards were not adopted for the DIT by EPAC, the analysis revealed that seniors and freshmen scored in the midrange identified by Rest (1979) as typical of college students and general adult groups.

Interpretation of the results is complicated by questions of validity of the ETS measures and by external factors, which appear to have significantly depressed the essay scores. These are addressed in detail in the technical report. Suffice it to say, it was concluded that the ETS tests did not contain a sufficient number of items to distinguish general academic aptitude from the specific generic intellectual skills. On the essay test, it appears that a longer period of time may have been needed for many students to do justice to the writing assignment. Also, informal interviews indicated that a number of students concentrated their efforts on organization and style and did not proofread their papers to correct misspelled words and grammatical mistakes, despite instructions to do so. Such errors significantly lowered the essay test scores.

Conclusions

Finkler's technical report was reviewed by EPAC early in the fall semester. Discussion addressed the possibility of conducting follow-up studies of writing skills. The prevailing view that emerged, however, was that such studies likely would shed little light on student writing skills. Although the APEP essay scores were somewhat disappointing, they were not altogether unexpected. And they did indicate significant improvement in seniors as compared with freshmen. Recognizing that responsibility for curricular change rests with the respective college faculties, EPAC decided to refer the report to the college educational policy committees for their review. EPAC also recommended that the colleges consider establishing

additional writing requirements for upper-division students for the purpose of maintaining the writing proficiency developed in the lower-division English composition program.

Thus APEP formally concluded at UNO. The project produced the first comprehensive assessment of the academic skills of graduating students since the institution became a part of the University of Nebraska. Profiles of student abilities were developed for each college, and these have been shared with the respective colleges. Hence, the two basic goals of the project were accomplished. In retrospect, a considerably longer test would have produced a richer and more reliable set of data. However, a solid, instructive beginning was made and can serve as the basis for further efforts.

Notes

[1] In 1980-81, the college of arts and sciences adopted a series of modifications in its general education requirements. In addition, an ad hoc Task Force on University Requirements was appointed to consider a variety of issues, including the appropriateness of adopting a campuswide set of graduation requirements. The task force developed such a proposal; however, only three of the six colleges offering undergraduate programs had approved it as of June 1983.

Written By

John T. Farr

Institutional Contacts

Otto Bauer, vice chancellor for academic affairs
John T. Farr, assistant vice chancellor for academic affairs

Wayne State College (NE)

Institutional Profile

Wayne State College, located in Wayne, Nebraska, a city of about 5,500, was established in 1909 when the state legislature authorized the purchase of the property, buildings, and equipment of the Nebraska Normal College, a private institution founded by Professor James M. Pile in 1891 and operated under his direction until his death in 1909. It opened as a State Normal School in September 1910.

In 1921, by act of the state legislature, the school became a State Normal School and Teachers College with legal authority to grant baccalaureate degrees in education. In 1949, the legislature changed the name of the institution to Nebraska State Teachers College at Wayne and granted authority to confer the baccalaureate degree for study in liberal arts. The graduate program leading to a master's degree was authorized in 1955, and in 1963 the legislature changed the name of the college to Wayne State College.

Wayne State College has an average enrollment of about 2,800 students on the Wayne campus during the regular session. Summer enrollment totals about 1,000 in the May, June, and July sessions.

Wayne State College is authorized to grant the following degrees: Bachelor of Arts in Education, Bachelor of Fine Arts in Education, Bachelor of Science in Education, Bachelor of Arts, Bachelor of Science, Master of Arts in Education, Master of Science in Education, and Education Specialist in School Administration.

Background

The phrase "global village," coined by Marshall MacLuhan more than a dozen years ago, emphasized the impact of media and technology on man's global perspectives and reinforced the perception of an escalating interdependence of mankind. As a result of almost-daily technological advances, today's students are moving into an enigmatic world of values confrontation, depersonalization, interpersonal communications anomalies, conformity, and ambivalence. Students' ability to put conflicting evidence into perspective, to sort out the meaningful from the trivial, and to discriminate between the enduring and the ephemeral will be crucial to their contribution toward helping to solve rather than becoming part of society's problems.

Institutions, through their general education programs, can promote skills to be used as lifelong resources by students as they are forced to come to grips with the problems manifested by an ever-changing, complex society. These skills (or "student outcomes," as they are sometimes called) may, by consensus, be viewed as institutional goals. Several issues are unavoidable and must be confronted by each institution's faculty: (1) What is the effect of our general education programs upon our students' lives? (2) How might this effectiveness be evaluated? (3) How might principled, nonthreatening, and academically defensible changes be promoted?

Although most in higher education agree that general education programs are institution-wide programs shared by the full faculty, many faculty members have come to feel disenfranchised from the general education decision-making process. This feeling of being "on the outside" may precipitate the factionalization of an institution's faculty, "turf protecting," and obstructionism. At the same time, those who feel "on the inside" of the general education decision-making scene sometimes tend to become intransigent whenever they perceive a threat to their general education domains.

In an attempt to confront the issue of assessing student outcomes in general education, several projects have been generated. The purposes of this report are to examine Wayne State College's role in the Academic Program Evaluation Program (APEP) and to suggest a process for arriving at institutional agreement about the goals of general education, to examine the process of generating conceptual definitions in key general education skill areas, to propose a process for operationalizing these conceptual definitions, and to describe procedures for gathering longitudinal general education data that may serve as a basis for curricular generalizations.

Implementation

Establishing Institutional Goals

There is disagreement about what constitutes evidence explaining changes in attitudes, values,

personal traits, and other characteristics that affect a student's college experience. That students do change as a result of their college experience is accepted. Establishing the reasons, by using an appropriate assessment instrument, is clearly another matter.

More than twenty years ago, Paul Dressel indicated there were at least two distinct approaches to evaluating general education programs.[1] One was to determine the skills of students at matriculation and graduation. Improvement in the respective skill areas under examination is thus ascribed to the particular program designed to satisfy those specific objectives. The obvious difficulty with this method is that the observed changes cannot logically be credited to the specific program because of the myriad extraneous and intervening variables that may have influenced the student during the college experience.

A second and more valid approach to program evaluation is to assess student outcomes--that is, what students are able to do, as evidenced by observable and measurable criteria--at graduation. These student outcomes (also called "performance criteria") can then be evaluated relative to the degree to which program goals are being achieved. If, for example, students do not demonstrate the desired outcomes at graduation, the curriculum can then be modified to remedy the deficiencies for succeeding classes.

Student outcomes, then, are the ends or objectives of the curricular program, whereas the instructional procedures within the components of the program are the means by which these objectives are achieved. Objectives can be constructed on a course-by-course, department-by-department, or generic skills basis. However, it must be emphasized that in the case of institution-wide programs such as general education, the larger the basis of consensus among the general faculty about specific student outcomes, the greater the likelihood of achieving the program's objectives. To this end, an institution-wide committee composed of representatives from each academic component contributing to the general education program (including student representation) was constituted for purposes of determining program objectives, performance criteria, and evaluation procedures to assess program effectiveness.

Generating Generic Skills

Once the appropriate committee was constituted, it began to function as a conduit representing its various constituencies' views and providing feedback on the particular outcomes that should be manifest in the graduating student. However, the levels of hierarchy extant in an institution, compounded by a cumbersome number of groups to be represented on the committee, made it difficult to obtain consensus on conceptual and operational definitions of generic student outcomes.

The Wayne State committee used the following approach: each committee member was to ascertain from the respective academic component those broad skills that are deemed essential and should be apparent in the graduating student. The committee then proceeded toward consensus on which of the skill areas should be addressed and subsequently, in a deductive manner, to those specific student outcomes mentioned above.

Generating Conceptual Definitions

Upon agreement regarding which generic skill areas will be addressed, the committee proceeded toward the actual generation of conceptual definitions. Each of the representatives on the committee was asked to look at how each of the five skill areas was being promoted in each of the respective academic divisions.

The next step for the committee was the translation of these conceptual definitions into an applicable format.

Operationalizing Conceptual Definitions

"Operationalizing conceptual definitions" means describing, within the parameters of agreed-on definitions, those essential academic abilities which must be demonstrable and measurable by the time of the student's graduation. It was agreed that these student outcomes or performance criteria should be stated in behavioral terms and should specify mediating conditions under which each behavior is to be carried out.[2]

As an intermediate step between the conceptual definitions and the behaviorally stated objectives, the committee chose to illustrate desired student outcomes in general terms. For example, the committee chosen to consider quantification skills as one of the generic skill areas stated general student outcomes in terms of what students must be *able to do* under certain conditions. For example, students will be asked to complete a quantification skills assessment instrument prepared by an appropriate faculty examining committee. They should be able to demonstrate an acceptable level of performance as evidenced by the solution of problems within a specified time frame calling on the abilities enumerated below:

- the ability to measure mass, magnitude, duration of time, number by category on appropriate scale, etc., using the appropriate instruments and units of measure

• the ability to apply computer capabilities to problemsolving (and so on).

From this intermediate step the committee proceeded to enlist the aid of an "appropriate faculty examining committee" to rewrite the aforementioned abilities in behavioral terms. After completion of the behaviorally stated student outcomes, the process of designing and/or adopting an assessment instrument began.

Measuring Student Outcomes

Over the years, many instruments have been designed to measure student achievement. For the most part, these tests measure subject-based knowledge and understanding and do not address such issues as whether students' scores reflect achievement in broad skill areas. More important, such standardized instruments do not necessarily address institution-specific or program-specific student outcomes. Thus, a number of testing instruments were evaluated by the committee. C. Robert Pace, in *Measuring Outcomes of College,* noted several landmark studies worthy of mention here.[3]

The Pennsylvania Study, undertaken in 1928, 1930, and 1932, involved administering to 4,500 graduating seniors from forty-five Pennsylvania colleges a twelve-hour achievement test battery. Sophomores were also tested in 1930 and were again tested as seniors in 1932 in an attempt to produce evidence in support of the obvious notion that the more extensively one has studied in a field, the more one knows about it. As expected, test results substantiated this claim as mean scores in every one of the forty-five colleges improved between sophomore and senior testing. Each college also gained on all or nearly all of the specific subjects measured by the test (except mathematics) regardless of the student's major field of study.

The Tests of General Education produced by the Graduate Record Office were a battery of examinations requiring eight hours to complete. Students were tested in general mathematics, physical sciences, biological sciences, social stuides, literature, arts, effectiveness of expression, and vocabulary. More than 1,000 students from sixteen colleges were tested as sophomores in 1946 and again as seniors in 1948. Mean scores of seniors in all subject areas were higher than those of the same students taking the test as sophomores, thus confirming the results of the earlier Pennsylvania study.

The Graduate Record Examination Area Tests were introduced in 1954 and were designed to test knowledge in the broad categories of social sciences, natural sciences, and humanities with a 3-3/4-hour test battery. These tests represented a significant departure from the Pennsylvania study and the tests of general education. Rather than testing for content knowledge, the tests evaluated students' reading, comprehension, and interpret retation. Again, test scores of 3,035 college seniors in 1954 confirmed the findings of earlier studies regarding students' scoring patterns in their fields of specialization.

The College Level Examination Program was designed to give students college credit for knowledge acquired outside the collegiate context. In order to gather normative data, the tests were given to a national sample of college sophomores and students enrolled in the United States Armed Forces Institute courses. Again, conclusive data were obtained illustrating that the level of student performance was directly related to the number of college courses completed.

In the Undergraduate Assessment Program, Area Tests and Field Tests were closely related to the area tests of the Graduate Record Examination and were initially administered to 47,000 college seniors from 211 colleges during the period 1969-1971. The UAP Area Tests related to students' "areas of interest" and measured achievement in humanities, natural sciences, or social sciences. The UAP Field Tests, on the other hand, measured student achievement in one or more of eleven content-based disciplines.

The Educational Testing Service recently revised the area tests based on data gathered by testing approximately 16,000 seniors in 105 colleges, 2,200 juniors in 57 colleges, 4,200 sophomores in 46 colleges, and 2,800 freshmen in 30 colleges during the period from 1976-1978. Significant differences in student achievement as a function of amount of completed coursework were obtained and proved to be compatible with the findings of all of the previously described studies.

The College Outcome Measures Project, sponsored by the American College Testing Program, is designed to measure communications skills, problem-solving skills, values-clarification skills, functioning within social institutions, using science and technology, and using the arts. Because of the abstract nature of the areas being tested, test materials required students to respond to film excerpts, tape discussions, photographs, music, charts and tables, magazine articles, etc. This battery was particularly time consuming and required six hours for students to complete.

The Wayne State Approach

In an attempt to provide a detailed plan of action that might be used by other institutions, there follows the five-stage approach that proved to be

effective at Wayne State College. This approach evolved out of the work of the Center Associates, the staff, and the consultants of the Resource Center for Planned Change of the American Association of State Colleges and Universities.

Stage I: Development of Conceptual Definitions

A. *Objective*: To develop and come to initial committee consensus on concise statements defining each of the five skill areas to be promoted, i.e., communicating, quantifying, analyzing, synthesizing, and valuing. These definitions should describe the five skills conceptually as well as operationally and should readily achieve institution-wide understanding.

B. *Rationale*: In order to ensure as wide a basis of consensus and support as possible, it is proposed that the college's faculty senate, futures planning committee, and division heads/deans, be consulted during all stages of the project and that a general education planning committee composed of several students and faculty members from each of the academic divisions be convened. The names of the faculty members should be proposed by each division in consultation with the appropriate constituencies. The names of the students should be proposed by the student senate also in consultation with the appropriate constituencies.

The representatives on this planning committee, as part of their task, will be asked to look at how each of the five skill areas is being promoted in each of the respective academic divisions.

It might also be appropriate that at some future juncture, three additional faculty members be assigned to critique the work of the committee and that, if possible, each committee member and the three additional faculty members be remunerated for contributions beyond duties normally considered part of their workload.

C. *Recommendations*:

1. That each of the respective divisions be allowed to determine the method (either election, appointment, consensus) by which its committee representative is selected.

2. That, as early as possible in the process, faculty members, administrators, students, and standing college committees be apprised of the goals and processes of the project and be continually updated on the project's progress.

3. That the process of working toward the project's goals be recognized as being as valuable as the potential products that might emerge.

4. That administrative intervention be minimized and the perception fostered that the primary administrative function is facilitative in nature.

5. That the general education program be recognized as an institution-wide program to inhibit the development of vested interest groups, turf protecting, and hidden agendas.

6. That all work of the committee be considered tentative and subject to revision by the full faculty.

Stage II: Identification of Observable Performances and Criteria

A. *Objective*: To identify and develop procedures that will elicit student performance indicating the presence of the generic skills and to establish criteria for judging the adequacy of student responses.

B. *Rationale*: This stage requires that levels of performance be established for each of the five skill areas. In order to do so, the committee, in consultation with its constituent faculty and administration, must determine the range of performance levels (or specify desired student outcomes) appropriate to academic standards for undergraduates at the college. There are several issues related to the establishment of levels of performance that must be confronted: (1) Is it feasible (or desirable) to establish minimum levels of performance in these skill areas? (2) Is it feasible (or desirable) to compare input measures to output measures in trying to establish the institution's academic impact upon its students? (3) What implications are extant regarding the possibility of faculty overload in attempting to determine whether and to what extent desired student outcomes have been achieved? (4) How can objectivity be achieved in evaluating student performance levels?

C. *Recommendations*:

1. That the committee recognize and call upon those faculty and staff members having relevant expertise for counsel in determining appropriate performance levels.

2. That the committee recognize the importance of consulting the faculty-at-large concerning the procedures by which student performance evidence be gathered, interpreted, and reported.

3. That, although standardized instruments may be used to assess the various skills, the primary concern of the project is to address institution-specific outcomes within the parameters of the college's general education program.

4. That the most efficient and effective mode of evidence gathering be employed while being mindful of the faculty's workload and day-to-day professional obligations to their classes, with minimal disruption of normal curricular routine, with the least threatening (to both students and faculty members) possible

process of data gathering, analysis, and reporting.

5. That the process employed to determine the extent to which desired student outcomes are achieved be academically defensible and easily understood by the faculty.

Stage III: Assessment of Students and Programs: Design and Implementation

A. *Objective*: To determine procedures for identifing which students at the college will be assessed, to determine the logistics of data collection, to come to committee consensus on the design of the assessment process.

B. *Rationale*: How can we find out whether, where, and to what extent undergraduates at our institution develop communication, quantification, synthesizing, analyzing, and valuing skills? This stage of the project will result in a determination of where, within the curriculum, these skills are being promoted by instructors and manifested by students. Stated in other terms, this stage will result in faculty members, administrators, and students determining what "evidence" will demonstrate that students have acquired, or are in the process of acquiring, each skill. We should look beyond the mere completion of courses and programs for such evidence and should consider such factors as course content, course structure, and course methodology in addition to employing innovative ways of examining existing institutional and program data.

C. *Recommendations*:

1. That each committee member submit a report resulting from examination of each respective division's contribution to the college's general education program in light of the committee-generated definitions of the five generic skills.

2. That the three consulting faculty members submit critiques on all committee work to date.

3. That the committee determine which specific courses in the general education program will be examined and the method by which students enrolled in these courses will be sampled.

4. That a retreat/workshop be scheduled for members of the committee, the three faculty critics, and the faculty senate to discuss the goals and procedures of the project.

5. That institution-wide, general education program statistics be gathered on

a) Enrollment (overall) in selected general education courses during the last three years, including the number of sections offered, total enrollment, student-credit-hour production, and average section size.

b) Enrollment (by semester) in each selected general education course during the last three academic years, including number of sections, total enrollment, average section size, and percentage of each division's general education contribution to the total institutional general education program.

c) Student-credit-hour production in the selected general education courses (by division of instruction) as a percentage of the total general education program semester-by-semester for the last three years.

d) Student-credit-hour production (by semester for the past three years) as a percentage of the total undergraduate production for selected general education courses. This information should include general education student-credit-hour production and general education as a percentage of total undergraduate student-credit-hour production.

e) Undergraduate student-credit-hour production (by division, semester-by-semester) for the last three years.

f) Enrollment (by division, semester by semester) for the last three years, including average section size, and general education student-credit-hour production as a percentage of the total institutional general education and as a percentage of the total undergraduate student-credit-hour production.

6. That subsequent to the first semester's final examination period, student performance evidence gathered from selected general education courses be prepared for analysis.

7. That during the following semester, expert committees of the faculty and student body be formed to examine and evaluate student performance evidence.

8. That during the following summer, student performance evidence be analyzed and reports of this evaluation submitted to the project coordinator for synthesis.

9. That as a result of the performance evidence analysis, course syllabi be examined voluntarily by each respective faculty member to minimize the dissonance between stated course objectives and actual course objectives achieved.

10. That additional questions be added to standardized course evaluation instrumenis (e.g., IDEA, Kansas State University) to generate data on general education course objectives.

State IV: Evaluation--Make Judgments Regarding Student and Program Performance

A. *Objective*: To evaluate the data on student performance and determine the worth of the level of student generic skill attainment in relation to the role and mission of the college.

B. *Rationale*: The assessment design developed by

the institution should define the parameters of the evaluation approach. The evaluation component of the project should seek to determine the extent to which the institution has promoted undergraduate competence in each of the five skill areas. In other words, based on the approach to assessment, faculty members should identify and perhaps design measures to demonstrate whether or how well the college has accomplished its goals for student achievement in each of the five skill areas. In this regard they should:

- specify where and to what extent in each academic division, each skill is being promoted
- specify appropriate measures for each skill area, ranging perhaps from unobtrusive measures to formal testing to expert judgment
- design and/or identify the necessary and appropriate administrative structure to carry out the evaluation plan.

C. *Recommendations*:

1. That committee members examine the general education curriculum on a course-by-course basis to determine which courses and constituent students would best be sampled to meet the project's goals.

2. That faculty members teaching general education courses be given the opportunity to volunteer to take part in the evaluation stage of the project.

3. That student performance data on a course-by-course basis be continuously gathered. This evidence may include copies of quizzes, tests, essays, final exams, projects, term papers, and so on.

4. That the anonymity of all students be preserved throughout the duration of the project.

5. That faculty members be assured they will not be judged by the performance of their students.

6. That very close contact be maintained between committee and faculty members teaching the general education courses being examined.

7. That during each semester, faculty members teaching general education courses be encouraged to update course syllabi.

8. That criteria be identified to determine the process by which the "expert" faculty evaluation committees (which will evaluate student evidence) will be formed.

9. That the "expert" faculty evaluation committees identify criteria by which student evidence will be evaluated.

10. That during the subsequent summer, the actual evaluation and analysis of student evidence be carried out, with committee reports submitted to the project coordinator.

State V: Policy Development and Decision Making

A. *Objective*: To remove discrepancies between desired and actual levels of student attainment; to reenter various project stages if definitions need modification or if new assessment/evaluation procedures are to be adopted.

B. *Rationale*: Decisions must be made after the analysis of student performance data regarding the remediation of any discrepancies between the real and the ideal levels of student performance in the various skill areas. Determinations must be made on whether such discrepancies are extant because of a lack of appropriate resources, an inappropriate allocation of available resources, an inappropriate curricular program, or an ineffective curricular design. In order to minimize these discrepancies, institutional policies and procedures must be examined and modified in light of student performance data by institutional consensus.

C. *Recommendations*:

1. That, as a result of the analysis of student performance data, the committee be central in the task of recommending procedures by which the findings of the project are to be presented to the college's various constituencies for discussion.

2. That the college's administration promulgate the perception that the general education program is an institution-wide program and rests in the hands of the entire faculty.

3. That the faculty of the institution be helped to recognize that curricular evaluation (and resultant modification) is a desirable phenomenon in higher education and a necessary element in curricular evolution toward increased effectiveness.

4. That new policies resulting in curricular change are the result of curricular evaluation and *not* the evaluation of specific individuals or divisions of instruction.

5. That a mechanism be employed for the continuing evaluation of the general education curriculum.

Conclusions

The results of the Wayne State efforts to focus on academic competencies in basic intellectual skills, as developed by the APEP project, are presented below. The section includes data on four skill areas: communication, quantification, analysis, and synthesis.

Communication Skills

Definition

Competence in communication, in a broad sense, involves all those procedures allowing an individual to express him/herself effectively. Within this broad

range two distinct, clearly related types of competence may be identified. The first is verbal competence, or the ability to speak and write clearly and effectively. The second is nonverbal systems of codification, which range from aesthetic symbols to gesture and digital language systems. It follows that effective communication is a form of human behavior that may be exhibited in a variety of media and contexts.

Desired Student Achievement

Students receiving the baccalaureate degree from Wayne State College should have the following abilities:

- to convey effectively information and ideas in speech, writing, and scientific and technical languages in a clear, coherent, correct, organized, concise, and appropriate manner
- to convey feelings, expressions, and values by means of verbal and nonverbal communication systems
- to receive and understand information, ideas, feelings, attitudes, and values from speech and writing
- to understand ideas, emotions, and values conveyed in art, literature, and music. (Expression through art, literature, and music is a significant form of communication. Although desirable to students achieving proficiency in such methods of communication, it is not something that can be required at the level of general education.)

Performance Level Criteria

Members of the communication skills subcommittee will evaluate the level of each student's performance on a scale ranging from highly effective to not effective. Highly effective performance indicates a superlative level of student achievement while performance judged not to be effective suggests that the student has serious difficulty displaying an acceptable level of achievement.

Data. Number of Course Evaluations = 34; Number of Different Evaluators = 4; Number of Student Evaluations = 868; and Overall Mean of Attained Communications Skills = 3.38.

Mean of Attained Communications Skills. General Education Courses = 3.41, n = 501 and Non-General Education Courses = 3.33, n = 367.

Mean Score of Attained Communications Skills by Course Level. 100 level = 3.38, Student n = 581; 200 level = 3.33, Student n = 254; 300 level = 3.76, Student n = 33; and 400 level = ----, Student n = 0.

Mean Score of Attained Communications Skills by Division. App. Sci. = 3.44, n = 317; HPERA = 3.23, n = 202; Business = ----, n = 0; Humanities = 3.39, n = 235; Ed/Psych = 3.90, n = 11; Math/Sci. = 2.97, n = 48; Fine Arts = ----, n = 0; and Soc. Sci. = 3.74, n = 55.

Mean Score of Attained Communications Skills by Each Evaluator. Evaluator I = 3.18, n = 312; Evaluator J = 3.52, n = 311; Evaluator K = 3.35, n = 234; and Evaluator L = 3.94, n = 11.

Quantification Skills

Definition

Just as college graduates must be able to express themselves in an organized, clear, concise manner using appropriate verbal and nonverbal symbols, they must also be familiar with basic quantitative principles and procedures.

Quantification is the ability to know when it is appropriate to measure, count, and compare and to know how to measure, count, and compare when appropriate.

Desired Student Achievement

Students receiving the baccalaureate degree from Wayne State College should have the following quantification skills:

- the ability to use appropriate instruments to measure number, mass, magnitude, and duration of time
- the ability to determine whether problems can be solved using quantitative techniques
- the ability to record data and to plot and interpret graphs and charts
- the ability to use calculators or computers to solve problems
- the ability to use arithmetic, geometric, algebraic, and statistical techniques to solve problems
- the ability to estimate quantities
- the ability to use basic formulas to convert units of measure
- the ability to determine probability
- the ability to use appropriate measurements in the analysis of art.

Performance Level Criteria

Members of the quantification skills subcommittee will evaluate the level of each student's performance on a scale ranging from highly effective (5) to not effective (1). Highly effective performance indicates a superlative level of student achievement while performance judged not to be effective suggests that the student has serious difficulty displaying an acceptable level of achievement.

Data. Number of Course Evaluations = 33 (some with combined multiple sections); Number of Different Evaluators = 3; Number of Student Evaluations = 1526; and Overall Mean of Attained Quantification Skills = 3.57.

Mean Score of Attained Quantification Skills. General Education Courses = 3.54, n = 1302 and Non-General Education Courses = 3.74, n = 224.

Mean Score of Attained Quantification Skills by Course Level. 100 level = 3.59, Student n = 1508; 200 level = 0, Student n = 0; 300 level = 2.00, Student n = 18; and 400 level = 0, Student n = 0.

Mean Score of Attained Quantification Skills by Division. App. Sci. = 3.35, n = 48; HPERA = ----, n = 0; Business = 3.62, n = 1334; Humanities = ----, n = 0; Ed/Psych = ----, n = 0; Math/Sci. = 3.22, n = 144; Fine Arts = ----, n = 0; and Soc. Sci. = ----, n = 0.

Mean Score of Attained Quantification Skills by Each Evaluator. Evaluator A = 3.56, n = 497; Evaluator B = 3.57, n = 507; and Evaluator C = 3.58, n = 522.

Analysis Skills

Definition

Analytical skill involves identifying the parts of a whole and the relationships of those parts to the whole. It includes the ability to identify the essential components of ideas, events, problems, and processes; to draw logical deductions about those components; and to recognize the limitations of these deductions. Because most problems are not isolated from one another, analytical skill includes the ability to establish relationships, when they exist, between problems.

Analytical competence also involves the ability to think critically and to solve problems.

Desired Student Achievement

Students receiving the baccalaureate degree from Wayne State College should have the following analytic skills:

- the ability to recognize what facts or assumptions are essential to or implicit in a thesis, claim, explanation, argument, or belief
- the ability to recognize and distinguish between the structural elements of oral or written communication, including introduction, premise, conclusion, and illustration
- the ability to devide compound or complex claims into simpler components
- the ability to detect relationships between ideas presented
- the ability to perceive connections between facts, events, or elements presented
- the ability to recognize how a theory, process, work of art, or scientific concept explains or orders diverse elements
- the ability to reason inductively and deductively
- the ability to assess the validity of inductive and deductive arguments
- the ability to detect cause and effect
- the ability to distinguish between the important and unimportant details in a passage or presentation
- the ability to detect faulty reasoning
- the ability to detect inconsistency and incoherence
- the ability to gather sufficient information to provide a data base for analysis
- the ability to distinguish between facts and assumptions.

Performance Level Criteria

Members of the analysis skills subcommittee will evaluate the level of each student's performance on a scale ranging from highly effective to not effective. Highly effective performance indicates a superlative level of student achievement while performance judged to be not effective suggests that the student has serious difficulty displaying an acceptable level of achievement.

Data. Number of Course Evaluations = 44; Number of Different Evaluators = 2; Number of Student Evaluations = 756; and Overall Mean of Attained Analysis Skills = 2.88.

Mean Score of Attained Analysis Skills. General Education Courses = 2.82, n = 476; and Non General Education Courses = 2.99, n = 280.

Mean Score of Attained Analysis Skills by Course Level. 100 level = 2.62, Student n = 498; 200 level = 3.09, Student n = 93; 300 level = 4.53, Student n = 19; 400 level = 2.73, Student n = 136; and 500 level = 3.80, Student n = 10.

Mean Score of Attained Analysis Skills by Division. App. Sci. = 1.59, n = 37; HPERA = 2.40, n = 115; Business = 3.00, n = 3; Humanities = 3.04, n = 237; Ed/Psych = 3.26, n = 90; Math/Sci. = 2.71, n = 56; Fine Arts = ----, n = 0; and Soc. Sci. = 2.89, n = 218.

Mean Score of Attained Analysis Skills by Each Evaluator. Evaluator D = 2.95, n = 409 and Evaluator E = 2.80, n = 347.

Synthesis Skills

Definition

Synthesis means putting together separate elements into a usable or satisfying whole such as a theory,

process, work of art, or physical construction. It is the organization of phenomena, events, ideas, or beliefs to yield a new order not previously apparent. (A theorist, for example, seeks to pull together seemingly disparate data into a comprehensible whole.)

Synthesizing skills are those which help an individual integrate basic knowledge, feelings, or the like. At all levels of every field of study and work this skill allows individuals to create innovative combinations of ideas, objects, processes, and people. Students with highly developed synthesizing skills are adept at putting apparently unrelated facts into a common frame of reference. They are able to find unifying themes or common threads that enable them to deal effectively with diffuse bodies of information. They not only possess a facility for constructing a general framework to accommodate whatever information they have, but they can move beyond that information, applying its principles to new situations or other bodies of information.

Desired Student Achievement

Students receiving a baccalaureate degree from Wayne State College should have the following synthesizing skills:

- the ability to organize related facts, ideas, and beliefs into a coherent framework
- the ability to introduce new facts, ideas, and/or beliefs into an existing framework
- the ability to extrapolate from known data to unknown
- the ability to recognize connections between theory and practical application
- the ability to use data and methods of different disciplines to understand a problem.

Performance Level Criteria

Members of the synthesis skills subcommittee will evaluate the level of each student's performance on a scale ranging from highly effective to not effective. Highly effective performance indicates a superlative level of student achievement while performance judged not to be effective suggests that the student has serious difficulty displaying an acceptable level of achievement.

Data. Number of Course Evaluations = 58; Number of Different Evaluators = 3; Number of Student Evaluations = 1247; and Overall Mean of Attained Synthesis Skills = 3.62.

Mean Score of Attained Synthesis Skills. General Education Courses = 3.79, n = 584; and Non-General Education Courses = 3.46, n = 663.

Mean Score of Attained Synthesis Skills by Course Level. 100 level = 3.59, Student n = 780; 200 level = 3.49, Student n = 225; 300 level = 4.01 Student n = 162; 400 level = 3.42, Student n = 74; and 500 level = 3.67, Student n = 6.

Mean Score of Attained Synthesis Skills by Division. App. Sci. = 4.18, n = 130; HPERA = 3.58, n = 109; Business = 3.11, n = 207; Humanities = 3.95, n = 216; Ed/Psych = 3.35, n = 188; Math/Sci. = 3.22, n = 95; Fine Arts = ----, n = 0; and Soc. Sci. = 3.79, n = 302.

Mean Score of Attained Synthesis Skills by Each Evaluator. Evaluator F = 3.58, n = 848; Evaluator G = 3.58, n = 335; and Evaluator H = 4.50, n = 64.

Notes

[1] Paul L. Dressel, "Methods of Evaluation and Research," *General Education*, Ed. Louis B. Mayhew (New York: Harper and Bros. Publishing, 1960), pp. 167-168.

[2] Paul D. Plowman, *Behavioral Objectives* (Chicago: Science Research Associates, 1971), p. xxvii.

[3] C. Robert Pace, *Measuring Outcomes of College* (San Francisco: Jossey Bass, 1979), pp. 10-47.

Written By

Bart Brady-Ciampa

Institutional Contact

Donald W. Whisenhunt, vice president

Western Carolina University (NC)

Institutional Profile

Western Carolina University (WCU), a state-supported coeducational institution established in 1889 and located in the scenic Appalachian mountain ranges of Cullowhee, North Carolina. The university consists of the main campus in Cullowhee and resident credit centers in Asheville and Cherokee. It includes six schools offering degrees at the bachelor's, master's, and education specialist levels.

Background

The meaning of general education was the question at the heart of the Academic Program Evaluation Project (APEP) from the outset. Western Carolina University had moved very rapidly in the late 1960s and early 1970s from the status of a small college known primarily as a teacher training institution to the position of a regional university with five undergraduate schools and a graduate school. As the curriculum developed to accommodate new or expanded fields of study, the general education requirements had become a subject of heated debate. At the time the APEP idea was proposed, a WCU faculty committee had just completed a two-year study of the general education program and had proposed an extensive revision. Having taken part in the AASCU forum in which the concept of the project evolved, the vice chancellor foresaw that APEP could significantly affect the reconsideration of general education.

The APEP design called for defining and measuring five intellectual skills; the WCU general education plan suggested ten content areas and seven skills as appropriate to a general education curriculum. WCU's miniproposal to AASCU to participate in APEP emphasized the value of such an evaluation process in clarifying issues and directions involved in redefining general education. It was thought that APEP participation would reinforce interest in developing course descriptions and requirements that included considerations of accountability and measurement of outcomes. The opportunity to engage in pre- and post-outcome measures while the redesign of general education was in progress appeared opportune for estimating the relative effectiveness of old and new approaches to course and program objectives. The project was seen as a way to develop a sound basis for describing objectives in behavioral terms as well as for learning how to develop measures that could assist the faculty in judging whether intended effects had been achieved. The project design also appeared to afford ways to identify more clearly what program elements might contribute to desired general education outcomes. Thus, despite the integrity of APEP and the entirely separate function of its steering committee, nearly every early decision relating to the project was colored by its inseparable relationship to the general education reassessment.

The fifteen-member APEP steering committee was selected by the vice chancellor for academic affairs, who served as project director, and the director of special academic programs, who was project coordinator. The active involvement of the chief academic officer of the university was seen as essential to giving the project high visibility and to assuring faculty participants of the project's high priority within the administration. Additionally, as an expert in the field of measurement and evaluation, the vice chancellor was uniquely suited by training and experience to provide strong leadership for the project. Another vital consideration was to name a coordinator with a record of success in conducting long-term, multifaceted projects. Finally, the whole group had to be task oriented and capable of sustained commitment.

Selection was designed to provide representation from all five undergraduate schools. Individuals were selected on the basis of their record of leadership and credibility among their peers. Although all of the committee, including two replacements, were among the busiest and most committed members of the faculty, all but one of the rest of the committee stayed with the project for two years. The decision to recognize and cope with the problems of overload inherent in selecting the busiest faculty members has proven positive, primarily because these individuals have infused the effects of the APEP process into the continuing, wide debate on implementation of a new general education program. The indirect impact was significant throughout the two years of the project.

For the first year of the project, both the director and coordinator sought ways to expand the membership of the steering committee, with the intention of achieving more widespread involvement

and, hopefully, better acceptance of the projected program evaluation. However, the committee feared that "going public" with APEP would exacerbate and confuse the already heated general education debate.

The proposed general education plan stated that "general education should . . . foster intellectual self-reliance and ability through extensive practice in reasoning and critical thinking; in experiencing the varied methods by which truth and knowledge are acquired and applied in organizing, analyzing, and synthesizing data, knowledge, and ideas; and in communicating the results thereof." The general education committee had said in its report that "no course is to be accepted for the program unless its plan includes adequate and evaluative exercises by each student of written communication in the English language, oral communication, the use of scientific method, the use of critical and analytical thinking, the use of logical reasoning, and the making of value judgments." Clearly, APEP and general education could reinforce each other. The question was how to maintain the distinctly separate character of APEP long enough to produce meaningful definitions and measures of intellectual skills while the controversy over the content and structure of general education remained the central issue on campus. Repeatedly, the committee reached the same conclusion: keep the APEP committee small, avoid publicity, and keep working. Ultimately, only a few individuals were added to the committee after initial appointments. Given the continued controversy over general education on campus and the considerable success of APEP in achieving its goals without arousing faculty opposition, the decision not to expand the committee and to maintain a low profile appears to have been a good one for our particular circumstances.

Implementation

Project Design

The project called for assessment of four skills, with a fifth skill--valuing--optional. The WCU committee agreed to delay consideration of the skill of valuing and to begin the task of defining analysis, synthesis, communication, and quantification. Initial discussions led to comments about the futility of reinventing the wheel and suggestions that the committee adopt or adapt existing definitions or hierarchies such as Bloom's taxonomy. But subsequent reviews of the literature proved more helpful in providing a background for discussion than a basis for agreement.

Management of the task required maintaining a careful balance between providing adequate background information and avoiding an information overload. Faculty members were asked to review key materials in the field, but for the most part the common understandings that had to be developed came from hearing presentations brought to the group by the project director, outside consultants, and faculty experts. Ultimately the committee split into subgroups, developed institution-specific definitions of the skills, and presented them to the full committee for debate, revision and approval, a process that occupied nearly five months. The time devoted to this stage of the project was well spent. During these months the committee gradually overcame initial reluctance to wrestle with the hard questions of what they actually meant by analysis, synthesis, communication, and quantification.

The major difficulty in defining analysis was in determining the appropriate sequential order of subskills and in clearly delineating at what point the skill of analysis merges with synthesis. Discussions of synthesis led to use of terms such as *chaos,* which in turn brought about lengthy consideration of the nature of order itself and the extent to which the committee believed synthesis is a process of perceiving as well as creating order. The definition of *communication* proved more difficult to contain than the others, as the term seemed to include receptive as well as expressive skills and, under examination, expanded rapidly with the incorporation of nonverbal languages, including such disparate forms as the arts and computer technology. Finally, in defining quantification the committee found that in addition to being a process for identifying and manipulating magnitudes and symbolic relationships, it, too, was a means of communication employing mathematical and graphic symbols.

Agreement came slowly, but with it came two very positive side effects. The committee expressed a sense of exhilaration at having spent time thinking and talking with colleagues about what education really meant. And the group achieved a breakthrough of understanding when, after months of frustrating, time-consuming effort that seemed doomed to failure, they realized that the process of defining skills and designing measures was far more important and useful than any product that might result from the project. That same realization rescued the work from low moments throughout the two and one-half years the committee worked together. At the end, the process had been significant both in individual professional development and in the multiple effects on campus of these faculty leaders.

As the committee began trying to develop per-

formance measures, early definitions that had seemed acceptable proved unworkable. The first of many loop-back cycles occurred as the committee rewrote skill definitions to include specific, measurable subskills.

Often the main obstacle to progress in the discussions was simply the lack of a commonly understood term. When the committee seemed destined for perpetual disagreement in its effort to work from definitions of generic skills to observable indicators of these skills, the missing term emerged as *enabling skills*. The subskills came quickly after that point. A similar semantic breakthrough resulted from the use of *indicator* in place of *performance measure* or *test item*. The committee also found it more reasonable to develop an "assessment design" than an "evaluation plan." The search for a common vocabulary that was not only clear but also comfortable was a continuing challenge to managers of the project.

In an effort to give the committee some kind of common background in evaluation technique, two faculty members with specialties in the field were asked to present a workshop for the group. The workshop proved helpful in supplying a common vocabulary. Even more helpful, the workshop presenters were added to the committee. At the end of the first semester, small task groups agreed to work during the summer selecting or developing appropriate measures of the intellectual skills the committee had defined.

The plan for the summer called for a review of commercial test instruments by two of the committee's experts on measurement with recommendation for use or adaptation. The materials were also to be employed as models for the committee to imitate in writing homemade indicators. Having agreed that the scope of the design would extend only to general education, the summer task group began creating a pool of test items and behavioral indicators, all of which were outlined according to subskill, content/context, cue to elicit the subskill, tools and conditions, and performance criteria. By the end of the summer, the group could present a selection of more than seventy-five possibilities to the steering committee.

The major issue to emerge from the summer's work was not a matter of what measures to employ or what sample to assess, but rather an issue that had been too quickly dismissed months earlier: the purpose of the project. It had first appeared that by taking part in the national project WCU was simply adopting the purpose of the national project as its own. An early statement of purpose required no more than twenty minutes of discussion. The committee would "assess the skill levels of students in four intellectual skills identified as common objectives of baccalaureate education. identify and describe program elements in general education that are probable contributors to acquisition of these skills, and on the basis of that data evaluate and possibly revise the general education program." It seemed simple; nonetheless, it proved to be far beyond the scope of the design that emerged at the end of the first year's work.

The question of purpose provided the major area of debate as the full committee resumed meeting in the fall. What could the project realistically hope to measure and to what extent could data be regarded as sufficiently reliable for making program and policy decisions? The committee began to believe that such causal connections between program elements, student intellectual behaviors, and test outcomes as might exist could not be clearly demonstrated.

The APEP paradigm that provided an outline for the project evolved and changed as ten institutions tested its effectiveness. The WCU project found that the paradigm had to be seen as iterative rather than linear or totally sequential. Loop-back cycles were frequent, beginning with Stage II, in which the task of devising applications for abstractions had given the project a firm base in reality. Definitions, statements of purpose, and long-range plans for policy and program change gave way to the major question of what could be observed or measured as an indicator of the generic skill and then confidently related to a course or program element. At that point the committee reached a major agreement, that the project design was to be formative, descriptive, and hypothesis-generating rather than summative, causative, and hypothesis-testing. The new purpose statement reflected that change:

> The purpose of the Academic Program Evaluation Project at Western Carolina University is to assess the skill levels of students in the four intellectual skills identified as common objectives of the general education portion of a baccalaureate education and to identify, describe, and increase awareness of program elements in general education that are probable contributors to acquisition of these skills. We intend to carry out such evaluations as are possible within the time and resource constraints of the project and, dependent upon the adequacy of the data, to make such policy decisions as seem justified. We expect, however, that it will be necessary to use an iterative approach and to employ more complex evaluative designs later to get at some of the intricate causal connections that WCU will want to consider.

undertaken as part of the new general education monitoring system be designed with advice and assistance from the APEP committee. The committee suggested that the first and most appropriate vehicle for providing that assistance would be a joint meeting with the "General Education Monitoring" (GEM) Committee.

The work of the APEP committee was also instrumental in gaining support for expanding the work of the Learning Resources Center (LRC), a unit established under another grant. The linkage between APEP and the work of the LRC has given support to the task of evaluating programmatic and student outcomes as well as the effort to link evaluation components into locally derived innovations. Certainly the success of APEP members in developing and field testing instruments of their own creation has supported the policy position that LRC should be provided with resources for encouraging faculty development of APEP-like measurement instruments.

Another interesting outcome of APEP is its reinforcement of an existing policy that supported provision of funds for this university to engage in exchange projects with other institutions. APEP, although only one of several current intercampus exchange activities that have borne considerable fruit, was different in that it involved so many other institutions over such an extended period of time. The success of this extended project has given considerable support to the argument for continuation and expansion of funding the interuniversity exchange efforts.

Several policy issues currently before the university will continue to feel, at least indirectly, the influences of APEP experience. For example, implementation of the content and distribution requirements of the new general education program is now underway. The issue of whether the distribution requirements should contain courses emphasizing analytical and synthesizing skills was settled affirmatively, and APEP must receive some of the credit. Similarly, the faculty has shown considerable interest in identifying resources for teaching such generic skills as analysis and synthesis. Continuing discussion about which methods might be best, which might be used in which disciplines, and whether indeed some of these skills can be taught in courses as currently offered argues for special attention to these issues and development of basic policies on resource allocation for examination of these and related issues.

Finally, and perhaps most important, APEP has stimulated renewed interest in program evaluation, has increased awareness of instrumentation issues in instructional evaluation efforts, and has increased sensitivity to the problems involved. At the same time the question of resource allocation and the issue of directions to be taken in internal evaluation efforts have been placed high on the agenda of both the administrative and faculty governance structures. At least one major policy decision has already been made as a direct consequence of the project. The office of the vice chancellor for academic affairs has adopted a long-range policy on evaluation calling for support of evaluating measurement instruments, whether locally developed or nationally available, and for implementation of a longitudinal study of students as they move through the new general education program. Comparisons between the new and old general education programs made possible by APEP's initial work will provide a valuable data base. Additionally, the initial evaluation design can be continuously reevaluated, leading eventually to sound and accepted bases for examining other academic issues sure to arise in the future. The end of APEP is only the beginning of academic program evaluation at Western Carolina University.

Written By

Marilyn Jody, Bruce Henderson, and Robert Pittman

Institutional Contact

Yvonne Phillips, interim vice chancellor for academic affairs

Western Kentucky University

Institutional Profile

Western Kentucky University (WKU), a publicly supported institution established in 1906 as a two-year state normal school, presently offers a comprehensive range of undergraduate and graduate programs within its four undergraduate colleges and graduate college. Total enrollment is nearly 13,000, with a full-time equivalent student enrollment of about 10,000. Of the students, 85 percent are undergraduates, nearly 85 percent are Kentucky residents, and more than 50 percent are women. The more than 650 faculty members represent thirty-six disciplines and specialties. The range of academic offerings includes technical programs at the associate and baccalaureate degree level; traditional baccalaureate programs in the arts and sciences, education, agriculture, business, health, and other broad fields; special professional and preprofessional curricula; and a number of graduate programs at the master's and specialist degree levels.

Background

Inclusion as one of ten AASCU institutions in the Academic Program Evaluation Project (APEP) provided WKU with a unique opportunity to examine even more closely the results of its instructional programs. Already a national leader in the evaluation of teacher education programs, WKU had more recently implemented a system for periodically reviewing all major programs. Neither of these efforts, however, assessed student attainment of the basic intellectual skills. For some time concern had been expressed among the faculty that students who enter without adequately developed skills in the area of written communication, quantification, and critical thinking are not provided ample opportunity within the baccalaureate program for developing each of those skills to the desired level and sustaining or improving them through continuous use. Thus, APEP was viewed as a means for beginning to examine that development.

APEP at Western Kentucky Univc sity was undertaken to evaluate student performance in four areas of intellectual competency--communication, analysis, synthesis, and quantification. A simulated pre- and post-test design was planned: a stratified (by college) random sample of graduating seniors and a group of incoming freshmen (matched to the seniors in terms of scores on the ACT mathematics and English scale) were to be tested in each of the competency areas and their scores compared. By using this procedure, the contribution of the program of study to the development of the intellectual skills was to be estimated.

The initial APEP organization consisted of the vice president for academic affairs, who served as project director; the associate vice president for academic affairs, project coordinator; and the general education committee for the academic council, which also functioned as the APEP committee during the first year of the project. This standing committee consists of fourteen members elected annually from the membership of the academic council: two faculty members from each of the four undergraduate colleges, two representatives from the academic services area, and two students. In addition, two administrators serve as ex-officio members of the committee.

Skills Definitions

The project director and project coordinator (an ex-officio member of the committee) held an initial meeting with the APEP committee to provide information, answer questions, identify the tasks to be accomplished, and discuss potential benefits of the project. The need for thorough understanding of the project and some initial reluctance to accept the challenge was exhibited during the first several of the weekly meetings. To assist in that regard, the project coordinator drafted a WKU-APEP "Focus and Goals" statement (end of chapter), which was approved by the committee and distributed for information and comment to the entire WKU faculty. Included in the statement were the following goals for WKU-APEP:

- to define each of the four intellectual skills reflect the achievement expected of all students earning a baccalaureate degree
- to specify for each skill area the average skill level attainment expected of students
- to ascertain where in the university, in what courses and experiences, students have the opportunity to acquire each of these intellectual skills

- to determine for each of the four skill areas the average skill level attained by randomly selected students
- to compare expected skill level attainment with actual skill level attainment in each of the four skill areas
- to make recommendations for change to the academic council if the evaluation indicates that student achievement in any skill area is less than that expected by the university
- to prepare a case study of the project experience.

In addition, the statement indicated that the project should result in improved methods for evaluating programs, increased efforts at program evaluation on the WKU campus, and improvement in the effectiveness of instruction.

Implementation

The initial charge of the committee was to define the skills. Conceptual definitions for the intellectual skills were formulated by four subcommittees, each with responsibility for defining a particular skill. Drafts of these definitions were submitted for review to the faculty-at-large and revised by the committee in response to criticism and suggestions. (The definitions that resulted are presented at the end of this chapter.)

Committee meetings continued weekly. Throughout the process of defining the skills, the committee sought the opinions of and guidance from the faculty. Although no attempt was made to hold a forum, interaction between the committee members and the general faculty was encouraged and was found to occur spontaneously within departments, colleges, and the university as a whole.

Assessment Instruments

Achievement of the first goal (the operationalized definitions) also marked the end of the academic year and the completion of the terms of general education committee service for the elected members. To ensure uninterrupted continuation of APEP, the project director and coordinator decided to form for each skill a new subcommittee, chaired when possible by an individual who had served on the original committee. Three of the four new subcommittees were chaired by faculty members with no previous involvement in APEP. To obtain a new committee, the college deans were asked to appoint to each subcommittee a faculty member who possessed both interest in the project and the appropriate expertise. Chaired by the WKU APEP coordinator, this committee was charged with reviewing available assessment instruments and selecting or developing instruments suitable for assessing the previously defined generic skills.

Because generic skills development should occur throughout the baccalaureate experience, the degree program should provide an appropriate scope for the assessment of seniors. However, incoming freshmen lack such experience, and a senior-freshmen comparison would not be invalid. Thus, a "life" scope assessment, involving situations that should be equally applicable to both groups, was chosen.

A necessary step before assessment instruments could be selected was identification of subskills, a difficult task in that four of the committee members who formulated the definitions had not been able to continue into this next phase of the project. One advantage of the change, however, was the opportunity to get more faculty members closely involved in the project.

One task of this stage of the APEP was not accomplished: the establishment of performance levels. Like the original APEP committee, the new subcommittees resisted establishment of expected performance levels. Some members believed that expectations of performance were inappropriate in the absence of the actual assessment instruments to be used; also many believed that previous performance data were necessary. Changing from a normative-referenced viewpoint to a criterion-referenced one seemed to be difficult for the latter group.

To accomplish the goal of determining at what points in the university learning experience the four skills are learned, the faculty was surveyed. The results identified the academic areas and the specific courses in which students are given the opportunity to develop the skills defined by the WKU APEP. With goals one and three attained, the groups moved to the assessment task.

The subcommittees decided that a test of two hour's duration for each of the four skills not only would be appropriate for the assessment but also was consistent with the extent of available resources. The main issue facing each subcommittee was the suitability of "home-grown" versus commercial instruments. The identified subskills and the content and scope of the assessment were compared with various commercial instruments to determine the extent of congruence.

The communication subcommittee determined that the written communication portions of COMP/ACT Activities 10, 11, and 12 appropriately addressed the subskills defined for communication.[1] This assessment requires a panel of six raters who must be trained utilizing materials supplied by

COMP/ACT. Analysis by COMP/ACT includes comparisons to norms of performance established by that company.

The quantification subcommittee decided on an assessment instrument composed of multiple-choice items and consisting of three portions--mathematical, statistical, and dimensional. For the mathematical thinking portion of the instrument, the subcommittee selected fifty-three items from an experimental design for assessment of mathematical thinking skills authored by Jonathan Warren of the Educational Testing Service (ETS).[2] The statistical and dimensional portions of the assessment instrument were locally derived. The subcommittee developed item pools, from which twenty-two items for each of the two portions were selected. Thus, the complete instrument for quantification consisted of ninety-seven items.

The analysis and synthesis subcommittees also decided that multiple-choice items would adequately address the skill definitions for analysis and for synthesis. Members of the analysis subcommittee developed a pool of locally derived items and selected twenty of the items from the ETS experimental design. In addition, they elected to purchase the Cornell Test of Critical Thinking, Form X.[3] Members of the synthesis group also developed a pool of locally derived items, from which forty-two items were selected. Additionally, they elected to use twenty-three items from forms 41, 42, 43, and 44 of the ETS experimental design.[4] Thus, sixty-five multiple-choice items constituted the total instrument for measuring the synthesis skill.

Concern for test reliability and validity affected instrument selection and design. Gathering data to establish the reliability and validity of locally developed tests is a time-consuming process, hence the decision to use commercial instruments and portions of the ETS materials as much as possible. Additionally, with the levels of performance expected of WKU students not established, the availability of performance norms was considered to be of great value.

Skill Assessment

A random sample of senior-level students, stratified by college, was selected from among students who (1) were scheduled for May 1981 graduation, (2) had matriculated at WKU as beginning freshmen, and (3) were completing degree requirements in five years or less. These criteria reduced the chances that student gains resulted from experience at another college or university or in a noncollegiate environment such as military service or full-time employment.

Incoming freshmen were chosen from among students who had attended the early orientation and registration sessions held during March and June 1981 in preparation for fall matriculation. Because the ACT scores of these students were available, it was possible to "match" a freshman to each graduating senior. The goal was to test at least fifty students from each population. To encourage the students' cooperation, they were offered free movie and bowling tickets, free meals, and (for the freshmen) free overnight accommodations.

The testing of each group, seniors during spring semester and incoming freshmen during the summer, was conducted over a two-day period with two of the skills evaluated each day. Examinations were supervised by the project coordinator and by subcommittee chairpersons. The four subcommittees worked independently to score the examinations and analyze the data. Abstracts of the subcommittees' reports, prepared by the chairs of those groups, follow.

Synthesis

The final 65-item form of the Synthesis Evaluation Instrument (SEI) was determined largely by the conceptual and operational definitions formulated by the General Education Advisory Committee, chaired by Phillip G. Duff, associate professor in the department of psychology. However, additional criteria for discipline freedom/fairness and face validity were consistently employed in final item selection.

Analyses of data generated by senior and freshmen samples revealed the following:

- *Reliability of the SEI.* When the two reliability coefficients were appropriately averaged, a mean reliability of internal consistency of 0.84 resulted--a value sufficiently high to warrant the adjudgment of significance.
- *Validity of the SEI.* When the correlation coefficients between SEI and ACT performances were appropriately averaged, a mean correlation of 0.67 was derived. The value of this coefficient can be interpreted as evidential support for validity of the SEI while at the same time militating against possible charges of redundancy.
- *Differences between SEI Performance Levels.* The most obvious--and distressing--difference between the performance levels of the two groups related to their respective means. Athough senior sample SEI mean performance exceeded that of the freshman sample, statistical analysis indicated insignificance of that difference.

• *Acceptable Level of Performance.* By utilizing the necessary chance performance statistics, the familiar "z" formula, and a desired level of confidence, a value reflecting statistically significant deviation from change-level performance was identified. This value was then established as the acceptable level of performance on the SEI.

Communication

The Communication Skill Assessment Instrument

The communication subcommittee, chaired by Mary Ellen Miller, associate professor, department of English, selected activities 10, 11, and 12 of the COMP/ACT as the assessment instrument for communication. These activities test for listening comprehension, ability to follow instructions, vocabulary, writing mechanics, organization, development, logic, and persuasion. An obvious advantage of the choice was the availability of norms, which ACT established by using data from seniors at twelve institutions that used Form IV of the COMP/ACT. Even so, the instrument was still considered experimental and under development.

Six faculty members of the WKU Department of English graded the tests. The graded tests were forwarded to ACT, where reviewers judged the grading to be consistent with the ACT scales except for one section, which was graded by ACT staff.

Analysis of Data, Communication Skill

In analyzing the data, ACT staff members coded the senior students by the college in which the major area of study was located and constructed a subsample of freshmen which was matched to the senior group on ACT scores and (as far as possible) on college.

The "average senior" in the small and self-selected subsample of seniors scored at the 39th percentile compared with seniors in the reference group. Compared with the same senior reference group, the freshmen sample scored at the 28th percentile. The matched subgroup of freshmen obtained a mean writing score at the 23rd percentile. Thus, if the samples had been representative, assessment results would have indicated gains of 11-16 percentile points occurred over the period from freshman to senior year.

As might be expected, both seniors and freshmen who reported themselves as being in the college of arts and humanities scored higher than the other college groups. When the seniors were stratified into grade-point average subgroups, no relationship between grade-point average and writing proficiency was evident.

Evaluation of Results and Recommendations

The faculty graders found that the writing performance of some of the seniors was below the level needed to pass the qualifying examination for English 102, required for all freshman students. Since all the seniors tested had presumably passed the qualifying examination, these results suggested that the writing skills of some students had degenerated between English 102 and graduation. The Communications Subcommittee recommended further testing of seniors to confirm or refute that finding. Consequently, the 102 exam was given to one or two classes of seniors in each college, with the grading being done by faculty who normally grade 102 papers. These faculty were not informed as to which papers were written by seniors. In this followup twenty-seven percent of the graduating seniors (compared with 10 percent of the freshmen) were found to have failed the exam. As a result of these findings, a task force on "Teaching Writing and Thinking Across the Curriculum" is now in action at WKU.

Analysis

The Analysis Skill Assessment Instrument

The instrument used consisted of form X of the Cornell Critical Thinking Test (CCT) and a twenty-item test developed by the Analysis Subcommittee, chaired by Larry O. Mayhew, associate professor, department of philosophy and religion) from the analysis scales 31, 32, and 33 provided by Jonathan Warren of ETS. The manual for the CCT suggests that the instrument is a valid test of the following aspects of critical thinking: detecting validity and invalidity of reasoning, locating assumptions, evaluating the reliability of observations and of authorities, generalizing from data, and assessing the relevance of reasons. The ETS scales include items that ask for general interpretation of information presented in a variety of modes (in written form, pictorially, graphically), for locating key assumptions and core issues, and for assessing logical inferences. Because the reliability coefficients of the ETS scales were low, the Analysis Subcommittee adopted eighteen items they deemed to be valid analysis items and completely rewrote two other items.

To gain more evidence of validity (especially for the ETS test), the tests were administered in late fall 1981 to students in an upper-level philosophy course. Before the results were scored, the instructor selected sixteen students whom he judged to be above average (in comparison with the general college population) in analytical ability. (The selection criterion was a grade of "B" or above on two course

papers that required analysis of journal articles in the philosophy of mind.) These sixteen students included eight graduate students from the M.A. program in humanities, specializing in philosophy, and eight undergraduate philosphy majors/minors (four seniors, three juniors, and one sophomore). The scores of these sixteen students were used for comparison.

Assuming that this group of students was above average in analytical ability, the results indicate that the instruments selected were able to detect differences in analytical ability. For example, on ETS the freshmen showed a range of scores from 3 to 19, the seniors from 5 to 19, and the (presumed) high group from 15 to 18. On the two instruments combined, the ranges were 34 to 77 for the freshmen, 30 to 83 for the seniors, and 64 to 80 for the high group. (The groups' respective means for CCT+ETS were 56.8, 62.2, and 71.5.) These results indicate that the tests are able to detect differences in analytical ability and provide an encouraging indication of their validity.

Correlation coefficients were computed for the instruments against ACT scores and (in the case of seniors) against grade-point averages. The moderate correlation between grade-point average and the instruments used indicated that the tests measured skills related to but different from academic ability. The moderate correlation between ETS and CCT reinforced intuitive judgments that these instruments emphasized somewhat different analytical skills. The large correlations between scores on the instruments used and ACT scores may have indicated that the skills measured are necessary for good performance on the ACT.

One important question about validity remains. If the group of philosophy students actually did possess a substantially higher-than-average analytical ability, is it possible that this group of students was only 8 percent (or even 16 percent) better than the average freshman at analyzing difficult issues? Obviously, important aspects of analytical ability are missed when only measured subskills are considered. The ability to apply the subskills collectively is probably an important aspect of analytical ability.

Evidence of reliability was obtained by using Cronbach's alpha, as a statistical measure of internal consistency. Reliability coefficients for the ETS were 0.79 and 0.73 and for the CCT were 0.85 and 0.81 for the seniors and freshmen, respectively. Thus, the subcommittee is "cautiously optimistic" about instrument reliability.

On each test, the mean score of the seniors was higher than the mean score of the freshmen, but the difference was not statistically significant. Given the available data, the members of the analysis subcommittee concluded that they were "unable to set standards and evaluate the performance" the seniors.

The setting of performance standards was the problem that seemed "most intractable" to the committee. Without them, no reliable means of interpretation was available. Although the committee considered three different ways of setting standards, serious objections were raised to each procedure. The committee recommended that the freshmen group be retested as seniors to explore changes that actually occur during the college experiences.

Quantification

The Quantification Skill Assessment Instrument

The evaluation instrument developed by the quantification subcommittee, chaired by John P. Russell, associate professor, department of industrial and engineering technology, consisted of three portions identified as mathematical thinking, statistical thinking, and dimensional thinking. The mathematical thinking portion, consisting of fifty-three multiple-choice items, was derived from an experimental design by the ETS. The items selected were those that demonstrated a biserial correlation with the overall test results greater than 0.45 and which fewer than 75 percent of the respondents answered correctly. These criteria were obtained from the results of a trial administration of the ETS design to approximately 1,500 college students at selected institutions. The statistical and dimensional portions of the instrument, each consisting of twenty-two items, were derived from item pools developed by the quantification subcommittee. The items were designed to be similar in scope and difficulty to those of the mathematical items, with the expectation that results from the three portions would be similar. The total instrument, therefore, consisted of ninty-seven multiple-choice items requiring approximately two hours to complete with no special materials required. When the instrument was administered to the group of seniors and to the freshmen, distribution of scores for the three individual portions of the instrument were identical, suggesting that they did provide equal discrimination among the students tested.

The quantification subcommittee chose a minimum performance standard of 50 percent correct responses. The decision was based primarily on the expected performance of college students on the ETS portion of the tests. Mr. Jonathan Warren of ETS had suggested that the "average college student" should score between 55 and 65 percent on the ETS items. Because

test results indicate that the three portions of the test were comparable in scope and level of difficulty, the 50-percent level seemed minimally acceptable for seniors.

Of the seniors tested, 23.8 percent did not perform at an acceptable level. The mean score of the seniors was 9 percent higher than the mean score of the incoming freshmen. The committee suggested, however, that "in order to describe a significance for the difference in test scores . . . some measure of sensitivity of the test to demonstrate a difference would be necessary."

When the test scores of the graduating seniors were matched with their composite ACT scores, a positive correlation between these factors was noted. The freshmen were tested in two groups. The fact that the mean test score of the second freshman group was 4 percent lower than the mean score of the first group also suggested such a correlation. Fifty percent of the total freshman group scored below the level deemed minimally acceptable for seniors.

Conclusions

Results of the WKU APEP were inconclusive. The inability to entice all of the selected seniors to participate and failure to establish performance standards for all four skill areas limit the conclusions that can be drawn. Several interesting implications can be made however. In each skill area, comparison of the scores of graduating seniors and the scores of incoming freshmen yielded a difference in mean scores for the two groups, with the seniors' scores being higher. The difference, however, was not significant in every case. Results also suggest that some of the graduating seniors are not performing at an acceptable level in all four skill areas.

The high correlation of the test scores with the students' ACT composite scores would seem to support the validity of the assessment instruments. Assuming that the ACT test measures attributes that are important in acquiring the generic skills, the instruments used in the assessment measure similar attributes or measure the acquisition of such skills (the latter was, of course, the intent).

The ability of the faculty to address cooperatively such broad concepts as the existence and measurement of generic skills without retreating into discipline-dominated stance is impressive. Faculty participants interacted freely and often commented on the benefits of that experience.

Finally, results of the WKU APEP, although inconclusive, provide a basis for further investigation of generic skill attainment. They suggest that the investigation of student outcomes can lead to further questions and that the search for answers to these questions can provide data that contribute significantly to the understanding of the educational profession.

Analysis

Conceptual Definition

Analysis is the process of hypothesizing the nature of the unified whole, identifying and classifying its component parts, determining the relation among parts, and recognizing the organizational principles involved.

Operational Definition

Analysis involves the following abilities:

- to hypothesize the nature of the unified whole
- to identify and classify the component parts so the unified whole

1. to identify the fundamental elements or component parts

a. to identify and distinguish among structural elements

b. to identify relevant details

2. to classify the essential features

a. to classify essential facts

b. to classify implicit assumptions

3. to elucidate and clarify expressions, statements, concepts, and principles

a. to infer the creator's purpose, point of view, or traits of thought and feeling expressed in his/her work

b. to infer the creator's concept of science, philosophy, history, or art

c. to describe the formal structure and expressive qualities of a work

- to determine the relations among parts of the unified whole

1. to detect the interrelations among ideas

2. to identify the information necessary to show cause and effect

3. to relate historical antecedents to subsequent or contemporary events

4. to comprehend the interactions among facts, concepts, events, and forces within a conceptual framework

5. to recognize ways in which method influences results

6. to identify the effects of motives, personal points of view, self-interests, biases, or prejudices on a communication or source of information

• to recognize the organizational principles involved in the creation of a unified whole
1. to identify the assumptions underlying the framework within which particular data are considered
2. to distinguish among hypotheses, theories, observation statements, normative statements, and definitions
3. to recognize how a theory, process, work of art, or scientific concept unifies, explains, or orders diverse elements.

Synthesis

Conceptual Definition

Synthesis is the process of organizing and integrating ideas, elements, or parts into a coherent whole--a pattern of structure which was not previously apparent.

Operational Definition

Synthesis involves the following abilities:
• to organize ideas, elements, or parts
1. to recognize basic differences, similarities, and other significant relations among items
a. using information and concepts available from current analysis
b. using information and concepts drawn from previous experience
2. to exclude those items not significantly related
3. to recognize possible organizational principles and to choose an appropriate one
4. to order the items in accordance with the organizational principle chosen
• to integrate ideas, elements, or parts
1. to choose and use existing organizational systems and concepts or to develop new combinations or organizational systems when necessary or appropriate
2. to develop generalized expressions as a means of explaining specific current phenomena.

Communication

Conceptual Definition

Communication is the process of conveying information, ideas, feelings, attitudes, and experiences to others and the process of receiving and understanding the information, ideas, feelings, attitudes, and experiences expressed by others, with the recognition that the communication may take a variety of forms.

Operational Definition

Communication involves the following abilities:
• to convey information, ideas, feelings, attitudes, and experiences to others
1. to select materials relevant to a particular problem or limited subject
2. to organize the selected material
3. to present the message clearly, logically, and concisely
4. to observe accepted conventions in the communication process
• to receive and understand information, ideas, feelings, attitudes and experiences
1. to identify the elements of a message received relevant to a particular problem or limited subject
2. to demonstrate an understanding of the message received.

Quantification

Conceptual Definition

Quantification is the process of determining and expressing the measured or relative value of an idea, a concept, or a physical entity.

Operational Definition

Quantification involves the ability to determine the measured or relative value.
• to recognize and understand standards of measurement
1. to know the units of measure appropriate to a particular situation
2. to understand the concept of number versus quantity
3. to choose appropriate entities or concepts to serve as basis for relative measure
• to apply measurement procedures and techniques
1. to recognize techniques applicable to a particular situation
2. to apply physical measures for determining quantity
3. to apply appropriate calculation techniques
• to recognize and apply statistical measures.

Focus and Goals Statement

Each recipient of a baccalaureate degree from Western Kentucky University is expected to have technical or professional competence in one or more fields of study. and each graduate is expected to have had a prescribed amount of academic course work in what is usually known as "general education." The definition of general education approved within the university expresses the belief that general education is a common academic experience designed to assist

students in developing wholesome self-concepts, an appreciation of their fellow man, and an understanding of the breadth and depth of accumulated knowledge. Attempting to ensure that students pursuing baccalaureate degrees will have the opportunity to attain the goals which are inherent in that definition, the university requires each student to complete a specific number of hours selected from approved courses within each of the following categories: organization and communication of ideas, humanities, social and behavioral studies, natural science-mathematics, physical development, and general electives.

Prior to this (1979-80) academic year, no effort has been made at WKU to evaluate the extent to which the meeting of these requirements in addition to "major" (and "minor" for some programs) requirements ensures students' attainment of specific intellectual competencies or of a specific level of knowledge in any one of the stated categories. In September 1979, Western Kentucky University became one of ten institutions chosen to participate in a 2-1/2-year evaluation project developed by the American Association of State Colleges and Universities and partially financed through the Fund for the Improvement of Postsecondary Education. The project's goal is to evaluate student attainment in four basic intellectual skill areas selected by the academic vice presidents of AASCU institutions. The chosen skill areas—are communication, quantification, analysis, and synthesis. At Western, the project is directed by James L. Davis, vice president for academic affairs, and coordinated by Faye Robinson, associate vice president for academic affairs.

Since the General Education Committee at Western has the responsibility for continuous purview over general education requirements, that committee (which is composed of faculty and student representatives from the academic units of the university) has been given the charge of working with the project director and coordinator to accomplish the initial tasks involved in evaluating the attainment by Western's baccalaureate students of the specified four intellectual skills. Persons from that committee and from the faculty at large will have extensive involvement as the assessment procedures are planned.

The goals of the Academic Program Evaluation Project at Western Kentucky University are:

- to define each of the four intellectual skills to reflect achievement expected of all students earning a baccalaureate degree
- to specify for each skill area to be evaluated the average skill level attainment expected of students
- to ascertain where in the university--in what courses and experiences--students have the opportunity to acquire each of these intellectual skills
- to determine for each of the four skill areas the average skill level attainment by randomly selected students
- to compare expected skill level attainment with actual skill level attainment in each of the four cognitive skill areas
- to make recommendations for change to the academic council if the evaluation indicates that student achievement in any skill area is at a level lower than that expected by the university
- to prepare a case study of the project experience.

It is expected that this project will result in improved methods for use in program evaluation, increased efforts at program evaluation on this campus and on the other campuses to which project information will be distributed, and improvement in the effectiveness of instruction.

Notes

[1] The American College Testing Program, P.O. Box 168, Iowa City, IA 52243.

[2] Jonathan R. Warren, Senior Research Scientist, Educational Testing Service, Berkeley Office, 1947 Center Street, Berkeley, CA 94711.

[3] Illinois Rational Thinking Project, 188 Educational Building, University of Illinois, Urbana, IL 61801.

Written By

A. Faye Robinson and John P. Russell

Institutional Contacts

Faye Robinson, associate vice president for academic affairs
John P. Russell, associate professor of industrial and engineering technology