DOCUMENT RESUME

ED 310 655 HE 022 716

| | |
|---|---|
| AUTHOR | Banta, Trudy W., Ed. |
| TITLE | Performance Funding in Higher Education: A Critical Analysis of Tennessee's Experience. |
| INSTITUTION | National Center for Higher Education Management Systems, Boulder, Colo. |
| SPONS AGENCY | Kellogg Foundation, Battle Creek, Mich. |
| PUB DATE | 86 |
| NOTE | 176p. |
| AVAILABLE FROM | NCHEMS Publications, P.O. Drawer P, Boulder, CO 80302-9752 ($12.95). |
| PUB TYPE | Reports - Evaluative/Feasibility (142) |
| EDRS PRICE | MF01/PC08 Plus Postage. |
| DESCRIPTORS | Academic Achievement; Accreditation (Institutions); *College Planning; Educational Assessment; Educational Finance; Educational Policy; *Educational Quality; Full State Funding; General Education; Higher Education; *Institutional Evaluation; Outcomes of Education; *Program Evaluation; *Resource Allocation; Self Evaluation (Groups); *State Aid; State Colleges; State Standards |
| IDENTIFIERS | *Performance Funding; *University of Tennessee Knoxville |

ABSTRACT

Changes at the University of Tennessee, Knoxville (UTK), begun when the Tennessee Higher Education Commission (THEC) initiated a funding mechanism designed to promote assessment and improvement of quality, are chronicled. Implications for program evaluation, planning, and resource allocation in other state colleges and universities are also discussed in the following chapters: (1) "Development of Performance Funding Criteria by the Tennessee Higher Education Commission: A Chronology and Evaluation" (Robert A. Levy); (2) "Planning and Resource Allocation at the University of Tennessee, Knoxville: Institutional Responses to Changes in State Funding Policies" (Homer S. Fisher); (3) "Comprehensive Program Evaluation at the University of Tennessee, Knoxville: A Response to Changes in State Funding Policy" (Trudy W. Banta); (4) "Accreditation as a Performance Indicator" (C. Warren Neel); (5) "Measuring Achievement in General Education" (W. Lee Humphreys); (6) "Measuring Achievement in the Major Field" (William H. Calhoun); (7) "Measuring Perceived Program Quality" (Kent D. Van Liere and William Lyons); (8) "Assessing the Quality of Higher Education through Comprehensive Program Review" (Mary P. Richards and C. W. Minkel); and (9) "Performance Funding and Institutional Response: Lessons from the Tennessee Experience" (Peter Ewell). Copies of the following are appended: THEC standards of performance, 1970; instructional evaluation variables and standards, 1981-82; instructional evaluation variables, November 21, 1983, THEC; instructional evaluation information for College D; student satisfaction survey excerpt from report for Department A; and guidelines for self-study document. References are provided with the individual chapters. (KM)

# PERFORMANCE FUNDING IN HIGHER EDUCATION: A CRITICAL ANALYSIS OF TENNESSEE'S EXPERIENCE

# PERFORMANCE FUNDING IN HIGHER EDUCATION. A CRITICAL ANALYSIS OF TENNESSEE'S EXPERIENCE

# Performance Funding in Higher Education: A Critical Analysis of Tennessee's Experience

Trudy W. Banta
Editor

With an Assessment by
Peter Ewell

1986

National Center for Higher Education Management Systems
P.O. Drawer P Boulder, Colorado 80302
*An Affirmative Action/Equal Opportunity Employer*

NCHEMS wishes to thank the W. K. Kellogg Foundation of Battle Creek, Michigan, for providing the funding support for the project on which this publication is based.

National Center for Higher Education Management Systems
Boulder, Colorado 80302

Printed in the United States of America
Production and design by Lynn E. Phillips

# Contents


Preface ..... vii

Introduction ..... 1

PART ONE: Performance Funding in Tennessee ..... 11

1. Development of Performance Funding Criteria by the Tennessee Higher Education Commission: A Chronology and Evaluation ..... 13
*Robert A. Levy*

2. Planning and Resource Allocation at the University of Tennessee, Knoxville: Institutional Responses to Changes in State Funding Policies ..... 27
*Homer S. Fisher*

3. Comprehensive Program Evaluation at the University of Tennessee, Knoxville: A Response to Changes in State Funding Policy ..... 37
*Trudy W. Banta*

PART TWO: Measurement Issues in Performance Funding ..... 51

4. Accreditation as a Performance Indicator ..... 53
*C. Warren Neel*

5. Measuring Achievement in General Education ..... 61
*W. Lee Humphreys*

6. Measuring Achievement in the Major Field ..... 73
*William H. Calhoun*

7. Measuring Perceived Program Quality ..... 85
*Kent D. Van Liere & William Lyons*

8. Assessing the Quality of Higher Education through Comprehensive Program Review ..... 95
*Mary P. Richards & C. W. Minkel*

PART THREE: An Assessment of Performance Funding ..... 103

9. Peformance Funding and Institutional Response: Lessons from the Tennessee Experience ..... 105
*Peter Ewell*

APPENDIXES ..... 121

A. THEC Standards of Performance—1979 ..... 123

B. Instructional Evaluation Variables and Standards—1981-82 ..... 129

C. Instructional Evaluation Variables—November 21, 1983, THEC ..... 139

D. Instructional Evaluation Information for College D ..... 153

E. Student Satisfaction Survey Excerpt from Report for Department A ..... 169

F. Guidelines for Self-Study Document ..... 173

# Preface

PUBLIC COLLEGES AND universities, like other institutions that are subject to regulation by external agencies, are periodically called upon to formulate policies that respond to changes in society. A growing trend is to use external authority—state authority—to promote high-quality education through assessment processes. Performance funding is an example. This book chronicles the changes set in motion at the University of Tennessee, Knoxville (UTK), when the Tennessee Higher Education Commission (THEC) initiated a funding mechanism designed to promote the assessment and improvement of quality in the state's institutions of higher education. From this case study, the authors draw implications for program evaluation, planning, and resource allocation in other state colleges and universities.

In Tennessee, up to 5 percent of each institution's annual instructional budget—a multimillion dollar figure for the largest universities—is awarded according to the criteria stipulated in the state's funding policy. This financial carrot has proven to be a powerful incentive. Tennessee's technical institutes, community colleges, and universities are now systematically assessing performance outcomes as part of program evaluation. This process has generated additional incentives for outcomes assessment, including the need to demonstrate institutional accountability, clarify the image of the institution, recruit and retain a diverse student population, enhance individual student development, improve program quality, carry out strategic planning, and allocate resources internally in ways that support the institution's mission.

Evidence collected in this book suggests that the use of performance criteria in the allocation of resources for higher education provides strong motivation for colleges and universities to initiate and strengthen program evaluation. The authors of this book describe an approach to comprehensive program evaluation, including measurement of student outcomes, that

can assist each institution of higher education when it addresses issues regarding quality in higher education. Because the cost in time and effort of faculty, students, and administrators that is needed to implement such an approach is substantial, one aim of this book is to demonstrate how the process can significantly benefit the institution. This book also aims to illustrate how the assessment of outcomes can help a college or university identify its strengths and weaknesses, thereby enabling it to clarify its mission and better plan its allocation of resources.

No other large research institution in the country has wrestled with a funding policy based on student performance like the one in effect in Tennessee. This dictated that 90 percent of the authors of this book were themselves participants in the events and processes that they have described. The possible subjectivity of such an evaluation is offset in three ways. First, the authors have diverse backgrounds and experiences. They include the chief administrative officer in business, planning, and finance, the vice-provost and dean of the Graduate School, the dean of the College of Business, and the head of the Department of Psychology, as well as professors in sociology, political science, and education. Second, most of the issues confronted and solutions proposed at UTK apply to other institutions of higher education. Third, Peter Ewell and the editors at the National Center for Higher Education Management Systems (NCHEMS) add an extrainstitutional perspective to this book that places the UTK experience in a national context.

This book is addressed to state legislators, staff and members of the policy boards of state coordinating agencies, college and university trustees, campus administrators, academic deans, and all others who are responsible for cultivating and raising the quality of higher education, particularly department heads and professors.

# Introduction

THE ERA OF GROWTH in higher education has reached a plateau. After two decades of expanding enrollment, with some lowering of academic standards as a by-product of this process, a primary concern in the 1980s is improvement of the quality of higher education (Bowen 1984, Folger 1984). Kenneth P. Mortimer, chair of the Study Group on the Conditions of Excellence in American Higher Education of the National Institute of Education (NIE), advocates that, because the era of high enrollments is now over, we "start thinking about advancing the quality of what we do" (Scully 1984, p. 20).

During the 1960s and early 1970s, at the height of enrollment growth and public recognition of the value of higher education, the flow of federal and state resources to a college or university was almost directly proportional to the increase in its enrollment. But in the 1980s, tax dollars are scarce, college degrees are less in demand, and public service consumers are demanding evidence of quality outcomes in return for investments in higher education. Colleges and universities are consequently being encouraged to define carefully the tasks in which they excel, focus resources on these endeavors, and evaluate and improve the extent to which they accomplish their objectives. Administrators and faculty members are being encouraged to develop evidence that dollars spent on a college education actually improve the student's preparation for employment, citizenship, and personal fulfillment (Edgerton 1984).

Past evaluation efforts have concentrated on measurements of resources such as expenditures per student, number of volumes in the library, percentage of faculty with a doctorate, and ability levels of students. New indices of the level of quality of higher education focus on student outcomes such as satisfaction with academic services, placement in employment, and achievement in general education and the major field. With declining enrollments and steady-state financing, competition among

institutions of higher education is increasing, as is students awareness that tuition dollars are best spent at institutions that offer educational experiences of the highest quality.

Today, in many states across the nation, there is increasing interest in performance funding. Assessments of quality in higher education by means of student outcomes is becoming an increasingly important consideration in the allocation of state funds to colleges and universities. Thus it behooves all institutions of higher education – public and private, large and small, two-year and four year – to pay close attention to student outcomes as a measure of quality.

## The Quest for Quality in Higher Education

"The typical college ... somehow manages to scrape along with little regular feedback relating to its mission. the education of the student" (Astin 1982, p. 14). The tradition that academia manage its own affairs was established in this country by the Dartmouth College decision in 1819 (Marcus, Leone, and Goldberg 1983). This precedent has shielded higher education from the type of probing by evaluators and accountability advocates that has become commonplace in public schools and social service agencies since the late 1960s and early 1970s. But today the quest for both quality and accountability in the American system, particularly in education, has led to public scrutiny of academia. The report of the NIE Study Group on the Conditions of Excellence in American Higher Education (1984), for example, notes:

> To assure excellence, our colleges, community colleges, and universities should establish and maintain high standards of student and institutional performance. The results (or "outcomes") of the education offered by these institutions must be measured against their clearly and publicly articulated standards of performance. [P. 3]

The NIE Study Group also notes that the responsibility for defining specific standards of content and levels of student performance and college-level learning in undergraduate education must fall on academic institutions themselves, or those standards will have no credibility" (p. 16). Finally, the NIE report recommends that:

> State officials ... establish special and alternative funding for both public and private institutions to encourage efforts that promote student involvement and institutional assessment. ... program improvement can inject an element of quality into the system that is not currently apparent. For public institutions, this funding might involve a set-aside percentage of the total state appropriations for public higher education. ... A few states have embarked on these efforts. [P. 70]

Additionally, the report of the Association of American Colleges (AAC) Project on Redefining the Meaning and Purpose of Baccalaureate Degrees, entitled *Integrity in the College Curriculum* (1985), points out.

> There must be ways of demonstrating to state legislatures, students, and the public at large that the colleges know what they are doing (or do not know) and that they are doing it well (or poorly). The colleges themselves must be held responsible for developing evaluations that the public can respect. [P. 33]

Thus, with respect to both fiscal and political concerns, it is increasingly the responsibility of administrators and faculty members to be able to demonstrate that their efforts are raising the quality of education that students receive today.

## Different Perspectives on Performance Funding

State agencies approach performance funding from a different perspective than do institutions of higher education. The state agency is interested in obtaining evidence of institutional accountability, eliminating program duplication, building regionally and nationally recognized centers of excellence through some differential allocation of resources, yet maintaining relatively convenient access to higher education for all citizens who desire and can profit from it. The state's purposes are best served if common evaluation standards are established that apply to all institutions and thus permit comparison of similar programs across institutions. Colleges and universities have some interest in demonstrating their accountability to the publics they serve, but they principally employ program evaluation to (1) determine which programs are central to the institution's mission, (2) identify the programs that should be maintained at current levels or phased out, and (3) discern how programs can be improved.

There are other differences as well. Colleges and universities prefer to establish evaluation standards based on their own unique missions and use measurement procedures that are valid for assessing their own programs. They prefer the richness of qualitative assessments that are organized according to local statements of objectives and priorities. Statewide standards are viewed with apprehension because they invite public comparisons of programs that are designed for different purposes, and because they can spark unhealthy and unwarranted competition among institutions. State agencies, on the other hand, prefer quantitative evaluation procedures that yield information that is easily handled, interpreted, and stored for future reference.

Despite the different perspectives on evaluation, public institutions have agreed over the years to supply state coordinating agencies with certain quantitative information from which implications about quality have been derived. Data on the proportion of the faculty with doctoral

degrees, credit hours generated, student/faculty ratio, expenditures per student, number of volumes in the library, entering ability level of students, external research dollars generated by the faculty, size of the institution's endowment, and so forth can be collected by each institution with minimal cost and disruption and, when compared across institutions, can provide an indication of institutional productivity and efficiency. Until recently, these indicators were also used to infer the quality of the educational experience afforded the students of a given institution. Increasingly, however, institutions are being called upon to furnish more direct, more valid evidence of product quality in the form of student outcomes. In other words, given certain levels of entering ability, resources per student, faculty preparation, and student/faculty ratio, what do students learn? How much value is added to their store of theoretical and practical knowledge as a result of the college experience? How well prepared are students for careers and success in a variety of adult roles? How satisfied are students with their experiences? Is there evidence that program faculties are engaged in continuous self-assessment for purposes of improving effectiveness and efficiency?

These are questions that can be answered only with new kinds of evidence, evidence derived from measuring student achievements and attitudes. This type of information is more costly, more intrusive, and more time-consuming to obtain than traditional data. Moreover, questions about instrument reliability and validity can threaten to undermine the whole endeavor. If state agencies and public institutions previously had difficulty joining hands across a narrow gorge representing disagreements over criteria for use in program evaluation, these parties now regard each other from opposite sides of a rift as large as the Grand Canyon!

### Performance Funding in Tennessee

Characterized by program architects Grady Bogue and Wayne Brown as a means "to improve the return on [Tennessee's] higher education investment" (1982, p. 123), performance funding was launched in Tennessee as the nation's first experiment in assessment of quality in higher education by means of student outcomes. "Acting on the possible while awaiting perfection" (Bogue 1980, p. 85) became the motto that was used by members of the Tennessee Higher Education Commission (THEC) to counter the many objections leveled against the use of measures of student outcomes in program evaluation. Having survived a rocky start and a barrage of criticism along the way, performance funding is now considered a qualified success in Tennessee.

In 1979, THEC instituted a funding policy based on performance criteria that has been applied to all publicly supported institutions of higher education in the state. This includes 7 regional universities, 10 community colleges, 4 technical institutes, and 2 comprehensive universities including the state's major research institution, the University of Tennessee,

Knoxville (UTK). UTK, by virtue of its size (25,000 students, including 5,500 graduate students), complexity, and research capabilities, has provided the setting for the most extensive field test of performance funding.

The THEC initiated a program that promotes and rewards systematic efforts to assess quality and improve instructional outcomes. Up to 5 percent of the instructional component of each institution's education and general budgetary allocation is awarded by the commission based on accomplishments with respect to five performance criteria. These include.

1. The percentage of programs eligible for accreditation that are, indeed, accredited.
2. The percentage of programs that have undergone peer review and/or have administered to majors a comprehensive field exam within a five-year period. Maximum credit for this standard is awarded if an exam is used and student performance improves over time or exceeds the performance of students in similar programs at comparable institutions.
3. Measurement of value added via the general-education component of the curriculum using the American College Testing Program (ACT) College Outcome Measures Project (COMP) exam. Maximum credit is awarded if the performance of seniors improves over time or exceeds that of seniors at a group of comparable institutions.
4. Measurement of opinion concerning the quality of academic programs and services using surveys of enrolled students, alumni, community members, and/or employers, and demonstrating that findings have been used to suggest specific improvements.
5. Implementation of a campuswide plan for instructional improvement, based on information derived from the procedures described above, as well as other sources.

The THEC has assigned a total of 100 points to these standards, with 25 points awarded for full accomplishment of the first, 30 points for the second, 25 for the third, and 10 each for the fourth and fifth (Tennessee Higher Education Commission, 1983).

In 1979, 2 percent of Tennessee's education and general budget for instruction was earmarked for program assessment and improvement. In 1983, the THEC was sufficiently convinced of the value and importance of its initiative in performance funding to increase the amount from 2 to 5 percent. The motivation for becoming involved is thus compelling. an incentive award equivalent to as much as 5 percent of each public institution's instructional budget is added to the institution's state allocation when an institution responds fully to the performance standards set forth in the Instructional Evaluation Schedule developed by the THEC. This amounted to $3.8 million in 1985 for UTK.

Changes in Tennessee's funding mechanisms incorporated at least partially some of the suggestions in a statement in the summary recommendations of the National Commission on Higher Education Issues:

> In addition to stable funding, the greatest contribution states can make to promote quality in the public institutions is to assure that their funding mechanisms are not overly enrollment-driven. Formulas based on enrollments provide little incentive to improve programs. Similarly, automatic year-end reversion of unexpended appropriations encourages spending and discourages both resource conservation and increased efficiency.
>
> State legislatures should avoid line-item budgets, position controls, and other mechanisms that work against the general principle of institutional flexibility within the framework of accountability. Such detailed budget requirements discourage effective and efficient administration. [1982, p. 4]

Tennessee's funding policy for higher education, in other words, has created a powerful incentive for university administrators to shift the focus of program evaluation from assessment of program resources to assessment of program outcomes. As a consequence of the state's new policy of performance funding, administrators and faculty members at Tennessee's colleges and universities have considered a variety of different ways to measure outcomes of instructional programs, and they have used the findings to suggest program improvements.

### Assessment of Performance Outcomes at UTK

In November 1981, UTK was selected as one of seven institutions of higher education across the country to participate in a grant program designed to increase the use of information about student outcomes. The grant program was funded by the W. K. Kellogg Foundation and administered by the National Center for Higher Education Management Systems (NCHEMS). Under the auspices of the Kellogg grant, UTK administrators, faculty, and students were brought together to (1) study the impact on campus programs of the THEC initiative in performance funding, (2) determine the most appropriate ways for a large, research-oriented university to meet the new performance criteria, and (3) consider how specifications about performance funding could be incorporated into and strengthen established institutional procedures.

Because the concept of performance funding originated outside the university and was first used by the state coordinating agency, it was initially viewed with skepticism by university faculty and administrators. Kellogg project monies were therefore used to find creative ways that would permit UTK faculty and administrators to feel that they owned the funding program. The Kellogg funds made it possible to capitalize on positive features of the funding criteria, that is, use them to supplement

and improve existing program evaluation, planning, and improvement processes. The measurement of outcomes and use of this information was thus incorporated into the university's academic program reviews, strategic planning, and internal allocation of resources.

Early in 1984, the Kellogg administrator for the project on student outcomes at NCHEMS nominated the UTK project for a triennial award given by the National Council on Measurement in Education (NCME). This award is for "an outstanding example of the application of educational measurement technology." Nominations for this competition are received from three sectors: private industry, the military, and education. Each of the three sectors is represented on the selection committee. In April of 1984, UTK was designated the recipient of the NCME award.

The conversations, research, recommendations, and implementation of recommendations that took place on the UTK campus in the interim between the awarding of the NCHEMS-Kellogg grant in November 1981 and the conferring of recognition by the NCME in April 1984 form the basis of this book.

## Design of This Book

This book is divided into three parts: (1) performance funding in Tennessee, (2) measurement issues in performance funding, and (3) an assessment of performance funding. In part one, performance funding is examined from the perspectives of state policy and university policies and programs, as well as from the perspective of administrators and faculty members. The authors analyze how changes in Tennessee's funding policy for higher education have brought about changes in planning and resource allocation at UTK, as well as provided an incentive for UTK administrators and faculty to evaluate existing programs. The authors discuss how criteria for performance funding were developed by THEC, the impact of performance funding on planning and resource allocation policies at UTK, and how performance funding has provided an incentive for UTK administrators and faculty to evaluate academic programs.

In chapter 1, Robert Levy, Associate Vice-President for Academic Affairs and representative for the University of Tennessee on the advisory body on performance funding throughout the 10-year development of the policy by THEC, sets out the history of performance funding in Tennessee. First he examines the initial pilot projects funded by Kellogg, Ford, and the Fund for the Improvement of Postsecondary Education (FIPSE), second, the interim drafts of the complex set of instructional evaluation variables that serve as performance standards for the program, and third, the adoption of the current evaluation schedule in late 1983. In chapter 2, Homer Fisher, Executive Vice-Chancellor and Chief Business Officer at UTK, analyzes how changes in Tennessee funding policies for higher education have engendered changes in planning and resource allocation at UTK. And, in chapter 3, Trudy Banta, Professor of Education and Director, UTK

NCHEMS-Kellogg Project, notes how the response to the policy of performance funding at UTK has been a complex evolutionary process with positive overall impact. She describes the instructional evaluation program fostered by performance funding at UTK, including its incorporation into program review, curriculum improvement, and institutional planning.

Part two of this book contains detailed accounts of UTK's response to the measurement issues raised by the state of Tennessee's approach to performance funding. Each chapter presents an analysis of one of the THEC evaluation variables according to the following: (1) measurement issues, (2) measurement methods used at UTK, and (3) methodological implications for assessment of program quality. These issues are addressed by UTK administrators and faculty members. In chapter 4, Warren Neel, Dean of the College of Business Administration, evaluates accreditation as a performance indicator. In chapter 5, measurement of achievement in general education is discussed by Lee Humphreys, Director of the Learning Research Center and Professor, Department of Religious Studies. In chapter 6, William Calhoun, Professor and Head, Department of Psychology, evaluates ways to measure achievement in the major field. In chapter 7, Kent Van Liere, Associate Professor, Department of Sociology, and William Lyons, Professor, Department of Political Science, discuss how to measure perceived program quality. And in chapter 8, Mary Richards, Associate Dean of the Graduate School and Associate Professor, Department of English, and C. W. Minkel, Vice-Provost and Dean of the Graduate School, evaluate ways to assess quality of higher education through program review.

Part three of this book is an assessment of performance funding in Tennessee by Peter Ewell, Senior Associate at NCHEMS and Director of the multi-institution, NCHEMS-Kellogg student outcomes project. Ewell examines the relevance of Tennessee's experience with performance funding for other institutions and state systems. He also notes national trends with respect to the issues of quality and accountability in higher education.

## References

Association of American Colleges [AAC]. *Integrity in the College Curriculum. A Report to the Academic Community.* Washington, D.C.: AAC, 1985.

Astin, Alexander W. "Why Not Try Some New Ways of Measuring Quality?" *Educational Record* 63 (Spring 1982):10-15.

Bogue, E. Grady. "State Agency Approaches to Academic Program Evaluation." In *Academic Program Evaluation,* pp. 69-87. New Directions for Institutional Research, no. 27. Edited by Eugene C. Craven. San Francisco: Jossey-Bass, 1980.

Bogue, E. Grady, and Brown, Wayne. "Performance Incentives for State Colleges." *Harvard Business Review* 60 (November/December 1982):123-28.

Bowen, Howard R. "What's Ahead for Higher Education? Opportunities for Optimism." *Change* 16 (April 1984):8-13.

Edgerton, Russell. "Abilities that Last a Lifetime. Alverno in Perspective." *AAHE Bulletin* 36 (February 1984):3-4.

Folger, John. "Assessment of Quality for Accountability." In *Financial Incentives for Academic Quality,* pp. 75-85. New Directions for Higher Education, no. 48. Edited by John Folger. San Francisco. Jossey-Bass, 1984.

Marcus, Laurence R.; Leone, Anita O.; and Goldberg, Edward D. *The Path to Excellence: Quality Assurance in Higher Education.* ASHE-ERIC/ Higher Education Research Report, no. 1. Washington, D.C.: Association for the Study of Higher Education [ASHE], 1983.

National Commission on Higher Education Issues. *To Strengthen Quality in Higher Education.* Summary Recommendations of the National Commission on Higher Education Issues. Washington, D.C.. National Commission on Higher Education Issues, 1982.

National Institute of Education [NIE]. *Involvement in Learning. Realizing the Potential of American Higher Education.* Study Group on the Conditions of Excellence in American Higher Education. Washington, D.C.: NIE, 1984.

Scully, Malcolm G. "National Concern over Educational Quality Seen Spreading from Schools to Colleges." *The Chronicle of Higher Education,* 12 September 1984, pp. 1,20.

Tennessee Higher Education Commission [THEC]. "Instructional Evaluation Variables." Nashville, Tennessee, 21 November 1983.

PART ONE

# Performance Funding in Tennessee

CHAPTER ONE

# Development of Performance Funding Criteria by the Tennessee Higher Education Commission: A Chronology and Evaluation

*By Robert A. Levy*

IN THE WINTER of 1975-1976, the Tennessee Higher Education Commission (THEC) initiated discussions to determine if a portion of the state's budget for higher education could be tied to performance criteria rather than to level of enrollment. Today, Tennessee educators continue to try to reach a consensus on ways to promote the assessment of instructional performance, and much progress in that direction is still occurring. This chapter begins with an examination of those discussions, it ends with a similar examination of related discussions still under way in 1984-1985. More specifically, this chapter addresses the following four issues: (1) the background and rationale for performance funding, (2) the design, implementation, and results of the funding initiative, (3) the prototype Instructional Evaluation Schedule, and (4) the current Instructional Evaluation Schedule.

## Background and Rationale for the Performance Funding Project

During the middle 1970s, publicly supported education institutions foresaw the end of the enrollment growth that had been triggered by the postwar baby boom and that had been used as a justification for ever-increasing state allocations. During this period, institutions also recognized that—with or without inflation—institutional expenditures would not level off in congruence with leveling enrollments. At the same time, the success of Howard Jarvis's Proposition 13 in California seemed to illustrate that the general population was growing more and more concerned about government spending, particularly in cases where good value could not be shown concretely for monies appropriated. Across the United States, policymakers and citizens were cultivating a value system based on quality rather

than quantity, and they were perceiving a need for large public entities to be more accountable to the population that supported them. It was in such an atmosphere that the THEC, a state coordinating agency for higher education with dual responsibilities of being an advocate for the postsecondary education community and a watchdog of that community, began to consider seriously how it might respond.

As part of a series of initiatives, the THEC elected to focus on possible changes in the funding formula for higher education. As in many other states, the THEC formula was designed to allocate appropriated dollars to the state's 10 community colleges and 11 universities in an objective and equitable fashion. Based on biennial audits of the cost of instruction, the THEC calculated the average cost of each credit hour by academic program (via taxonomies created by the Higher Education General Information Survey [HEGIS]) according to the following seven instructional levels: freshman-sophomore, junior-senior, master's, professional-law, doctoral, remedial education, and continuing education. These historical data were then used to build institutional budgets by, first, estimating the number of credit hours to be produced by each academic area and, second, multiplying this figure by the average dollar value at each level. (Note: This basic formula underwent annual revisions until the summer of 1984, at which time it was discarded in favor of a "peer institution" approach. Despite this major shift in philosophy, the Instructional Evaluation Schedule has been retained.) Although the formula had many subparts, it relied principally on this cost-averaging methodology.

Many institutional administrators claimed that the methodology had two main failings. One was the lack of quality assessments, the other was that it insidiously encouraged the average rather than the excellent. As a response to this criticism of the formula and as a response to the broader context noted above, the THEC initiated the Performance Funding Project. The January 1976 THEC publication *Allocation of State Funds on a Performance Criterion (Performance Funding Project)* stated, "The major purpose of the project is to explore the feasibility of allocating some portion of state funds on a performance criterion (how effective), as compared to the current allocation on activity criterion (how much)" (Bogue, Harris, and Troutt 1976, p. 23). A survey of the extant literature yielded little information oriented toward application, and in terms of empirical data, contacts with other state coordinating agencies and university systems proved similarly fruitless. In-depth conversations at each public institution and with the staffs of the state's two governing boards indicated a willingness to explore the concept further and, at the same time, demonstrated that the task would not be an easy one. Contacts with the American College Testing Program (ACT), the Educational Testing Service (ETS), and the National Center for Higher Education Management Systems (NCHEMS) showed that each of these agencies had some potentially useful research and development activities under way. The research was not, however, ready to

be tested, and it did not consider the relationship between assessment of quality and allocation of state resources.

A number of nationally prominent educators were invited to serve as consultants in connection with the project, and conversations with them resulted in a number of "blue-sky" ideas that were in need of refinement and testing. The THEC also established a broad-based, statewide advisory committee of educators, fiscal experts, and key legislators to help plan how the project should proceed. All of these activities, together with the formation of a national advisory panel and the writing of several project proposals to granting agencies, resulted in a general design for the project.

Before the outlines of the Performance Funding Project could be drawn precisely, it was necessary that all parties involved reach a reasonable degree of consensus on operating assumptions. In other words, the Performance Funding Project itself, the activities associated with it, and its possible outcomes had to be carefully considered because they had to be politically acceptable, academically respectable, fiscally sound, and flexibly responsive to the needs of the public institutions in the state.

To be politically acceptable, particularly at a time when state budgets were already stretched to the limit, the project had to be externally funded. In addition to a grant from the Fund for the Improvement of Postsecondary Education (FIPSE), monies were obtained from the Kellogg Foundation, the Ford Foundation, and a foundation that wished to remain anonymous. The project had to be politically realistic in a number of other ways, as well. First, results of the project had to be acceptable to the Tennessee General Assembly and its constituents, as well as to the state's community colleges and universities. Second, the project had to result in a funding process that would not necessitate significant additions to either the legislated responsibilities of the THEC or to its staff and budget. Third, the project had to be sensitive to a primary public policy question. If quality can be measured and if the state's resources are limited, then should higher quality institutions and programs be rewarded for their success, or should lower quality institutions and programs receive increased funding in order to make improvements?

To be academically respectable, the project had to involve faculty members and institutional administrators, and it had to employ a method that was at least quasi scientific in its accumulation and use of data. The results of the project had to mesh with existing fiscal policies and procedures, accounting/auditing systems, budget-making timetables, and the existing formula. Project results also had to allow for the development of new policies and procedures that would be acceptable to all the parties involved.

Finally, the Performance Funding Project had to take into account the variety of existing institutions in the state. Tennessee's public institutions of higher education were, and still are, operated by two wholly separate governing structures, namely, the University of Tennessee, headed by

its Board of Trustees, and the State University and Community College System, headed by a State Board of Regents. The state's public institutions included 2 large comprehensive universities, 8 smaller universities, a comprehensive center for the health sciences, and 10 community colleges. Each of these institutions had been encouraged to adopt unique institutional missions, coordinated by the THEC; each, in fact, took justifiable pride in defining its particular educational role and scope of activities.

All of the operating assumptions of the Peformance Funding Project, as well as the myriad questions and assumptions that they in turn generated, were discussed, drafted, and redrafted many times during the winter of 1975-1976. Although disagreements remained, at least some of the reasonably specific areas of agreement were identified, and by the spring of 1976, dialogue had progressed to the point that a design for the Performance Funding Project could be discerned.

## Design, Implementation, and Results of the Performance Funding Project

The design of the Performance Funding Project made it possible for several related activities to take place simultaneously, in practice, however, institutional pilot projects soon drew time and attention away from other activities. Among the other activities was the writing of a discussion paper called *The Competent College Student* (Branscomb, Milton, Richardson, and Spivey 1977) by four of the elder statesmen of Tennessee's educational establishment. This essay is about the knowledge and skills that, according to the authors, every college graduate should possess. It remains interesting even today; but in the ferment of activities then surrounding performance funding, the discussion of competence quickly centered on the issues of institutional autonomy and measurement accuracy.

By July 1976, each institution had been invited to submit a pilot project proposal. Each of the two governing boards was asked to appoint a liaison officer and invited to submit a pilot project proposal. The *THEC Guide for Proposal Preparation* (Bogue and Troutt 1976) outlined the following key elements to be included. (1) in order to protect the validity of the project and address aspects of the operating assumptions, each proposal "should outline a set of performance indicators reflecting the identity of an institution"; (2) the performance indicators should be related to the statewide goals of "improved instructional effectiveness, improved quality instructional environments, and improved quality instructional productivity"; and (3) each proposal "should provide at least some very tentative thinking about how performance on indicators might be rewarded through the appropriations process." For each pilot project approved, a one-year contract was executed, with a second-year renewal option. Proposals were submitted by 19 of the 21 public institutions, and 12 were eventually approved by the THEC. The THEC hired two independent, out-of-state consultants to advise its staff and pilot project directors and serve as

evaluators of the projects. A brief look at some of the pilot projects helps to explain the origins and background of the modifications the THEC eventually made in its funding formula.

Memphis State University (MSU) and the University of Tennessee, Knoxville (UTK), each undertook a pilot project. The two projects were markedly different. MSU examined the aspects of its curriculum in which all students would be involved, UTK studied activities within its College of Engineering, with the hope that the study's methodology might serve as a model for other collegiate units (or at least for other units with a professional orientation). Using the Institutional Goals Inventory developed by ETS, MSU attempted to examine three campuswide curricular goals. Two of the goals—communications and computation—were assessed by administering instruments provided by ACT. Preexisting ACT verbal and math scales submitted by students applying to MSU were used as a kind of imputed pretest, and they formed a base against which to judge post-test ACT scores of a sample of MSU seniors. More important for the future, however, was that MSU participated in the field testing of the ACT College Outcome Measures Project (COMP).

The quality of vocational preparation at MSU, the third goal to be assessed, was measured in two different ways. The first was via a locally developed survey of employer satisfaction, which measured student satisfaction with courses, programs, counseling, and advising services. The second was via field-specific standardized examinations, including the Engineer-in-Training (EIT) examination, the National Teacher Examinations, and the National League for Nursing examination. The EIT examination was also used in the UTK project in the College of Engineering. Additionally, in order to gather graduates' assessments of the quality of their education and survey employers' perceptions of the quality of career preparation of engineering programs, a survey was conducted of the nearly 700 engineering baccalaureate graduates of the previous 10 years and their current supervisors. The UTK project also called for the formation of a 23-member lay Board of Visitors to formulate collegewide goals and assess progress in meeting those goals.

Several of the participating institutions attempted to use standardized examinations of established validity and reliability. The University of Tennessee at Martin employed the Nelson-Denny Reading Test and the Rokeach Value Survey, the University of Tennessee at Chattanooga used the Watson-Glaser Test of Critical Thinking, the Concept Mastery Test, and the Strong-Campbell Interest Inventory, Columbia State Community College used items from the National Assessment of Educational Progress, and Middle Tennessee State University used the Stanford Diagnostic Reading Test. These instruments were typically administered in a pretest/post-test mode, and the brief duration of the Performance Funding Project necessitated the construction of inferentially matched cohorts. The real difficulty with such examinations, however, was the problem of relating

test scores to intended outcomes of higher education. Thus, a general result of the pilot projects was to affirm the legitimate, but limited, purpose of professional licensure examinations for certifying competence of students graduating from programs that were clearly directed toward a specific occupation or cluster of occupations, while at the same time demonstrating that extant, off-the-shelf instruments would not serve as valid indicators of the nonprofessional education component of curricula.

Some institutions elected to use the now-defunct Undergraduate Assessment Program (UAP) provided by ETS. The UAP test battery employed items from retired Graduate Record Examinations (GRE) in an attempt to assess both competencies in general education and proficiencies of college seniors in selected fields. The UAP did not, however, establish an acceptable level of validity for a test of general education because item analyses of its questions suggested that UAP/GRE questions were designed to rank individual examinees rather than assess the competence of test takers.

The pilot projects at Volunteer State Community College and Middle Tennessee State University field tested an instrument designed to assess competence in general education (the instrument was being developed by ETS but had never evolved past the research stage). Other pilot projects employed local tests of general education, but the difficulty of constructing a valid and reliable test proved to be unmanageable in the time available. The primary value was that it helped focus faculty attention on the place of general education in the curriculum.

At the end of a meeting in October 1976 at which institutional pilot project directors summarized their assessment activities (with emphasis on the difficulties they were encountering), the two external consultants had started to crystallize their thinking about the direction of the Performance Funding Project. One consultant noted, "We should be ever mindful that no single indicator is really defensible at all, and that even a battery of indicators (the only acceptable method) is far from perfect, nevertheless, the battery approach is vastly superior to nothing at all—e.g., the formulaic method in use in Tennessee and many other states" (Harcleroad 1976). At the close of that meeting, there was consensus among project directors on the following three points:

1. Compared to nationally standardized measurement instruments, instruments that are developed locally and instruments that are valid for the information content of the courses that are taught at a particular college or university are more relevant for assessing performance. Instruments that are related to programs or instruments that are linked to performance funding criteria are most effective. In other words, assessment of a student's performance should be as closely related as possible to the student's intentions, and use of indirect measures of student achievement (satisfaction

indices, employment surveys) should be avoided whenever possible. Assessments of student values and effective skills are even more problematic than those of general education and specific field competencies.

2. Competence levels and value-added measures should be considered jointly in order to (1) prevent more selective programs from being judged solely in terms of gain scores (because the gain scores might be lower than gain scores of other programs) and (2) give less selective programs the opportunity to demonstrate their instructional effectiveness (via value-added scores) without creating invidious comparisons with the exit scores of more selective programs (Levy 1976).
3. Institutions should motivate students to perform their best when being assessed so that the results are as credible as possible when compared with traditional measures such as grade-point averages, scores on admissions tests, and licensing examinations.

The various institutional pilot projects proceeded apace. There were regular meetings of all the project directors. Site visits were made by the THEC consultants as well as by other specialists and consultants. A variety of mid-course corrections were made, based on mid-year reports. By the summer of 1977, the THEC consultants were generally pleased with the progress that had been made. They noted:

> If we look for significant impacts in terms of activities and on-going events taking place on their respective campuses, we readily receive the impression that at least 7 of the 10 projects are making good progress toward their stated objectives. . . . This may suggest that the major outcome of the overall project will be the experience gained by project participants and the implications of that experience for continued institutional development. . . . We need not anticipate the success of the Performance Funding Project to speculate that the activities of planning, goal-setting, assessing, and evaluating are on-going functions of an institution that should lend themselves to incorporation in a budget formula. . . . Potential performance indicators may still be focusing too exclusively on student performance in terms of behavioral changes, career preparations, academic and job competencies, etc. . . . Where indices of faculty productivity, community or public service, and the uses of knowledge can be considered, they should be. . . . The trained professional will be the major outcome of many educational programs—but not the only product. [Fincher 1977, p. 1-5]

As the pilot projects moved toward developing their final reports in the spring of 1978, the effects of testing the operating assumptions became clear. It had been asked that the proposals be politically acceptable,

academically respectable, fiscally sound, and respectful of varying institutional missions. No longer was attention being given to the notion of statewide performance standards to which dollars would be linked according to an institution's level of performance. The THEC recognized that this initial model of performance funding would not be a realistic political possibility. Instead, an increasing amount of attention was paid to providing additional funds for institutions willing to make evaluation efforts and use a battery of indicators. Thus, at least initially, monies could be allocated on a basis other than actual level of performance.

Although many of the pilot projects' formal reports submitted in July 1978 included potential models for revising the THEC funding formula, none was deemed politically or fiscally viable. An alternative plan, adopted by the THEC on August 28, 1978, would have inserted some performance criteria into the formula by means of a special appropriation. The idea, and the dollars tied to it, were not included by the new governor in his 1979-1980 appropriations recommendation.

### Prototype Instructional Evaluation Schedule

Nearly four years had passed since the first open discussions about linking instructional results and state resources. During those years, a number of bright and creative people had discovered, explored, and discarded a frustratingly wide array of ideas, strategies, and instruments. While it was increasingly clear that there were ways to assess instructional outcomes, it was also increasingly clear that even the use of a battery of assessment devices was so recognizably imperfect that educators mistrusted it and politicans were leery of funding it. Virtually everyone associated with the Performance Funding Project had learned to deal with adversity by subscribing to what had become the project's motto. "Act on the possible while awaiting perfection."

The THEC staff developed a revised funding concept that was adopted in October 1979 and implemented in the fall appropriations cycle of that year. Since the Performance Funding Project had officially ended and since the THEC funding formula is a series of schedules, the new policy was called the Instructional Evaluation Schedule. Appendix A, the original schedule as approved and implemented, illustrates the lessons learned from the Performance Funding Project:

1. As often as possible, existing instruments were utilized, so as to continue to test the legitimacy of such instruments for these specific assessment purposes and to allow institutions to begin assessments in a timely way. An interesting corollary of this principle was the use of accreditation in the schedule as a performance measure. In this context, accreditation represented off-the-shelf procedures. However, it was generally agreed that accreditation tended to focus on inputs rather than outcomes.

2. The prevailing lack of confidence in any single measure of instructional effectiveness was illustrated by the use of six separate variables that were weighted equally.
3. Flexibility and increased acceptance by institutions were sought by the addition of the last variable, the intent of which was to recognize any and all activities that did not readily fit into any of the other variables. Moreover, since point calculations were based on scores of only the top five variables, each institution enjoyed a protective cushion.
4. In order to allay fears about unfair institutional competition, reinforce the importance of institutional mission, and stress the importance of the sequence of evaluation/action/improvement, the schedule incorporated both value-added features internal to an institution and the possibility of assessment indicators with regional or national reference standards.

To fund the Instructional Evaluation Schedule, the THEC added approximately two million dollars to its statewide appropriation request with the idea that the schedule would allow each institution to earn up to two percent of the instructional component of its education and general budget. (In a time of limited resources, 2 percent was large enough to command some attention from the institutions, yet small enough not to cripple an institution that scored badly or chose not to complete the schedule.)

At the appropriations hearings during the early fall of 1979, the Instructional Evaluation Schedule was not well received by the state's institutions of higher education. The natural tensions between the institutions and the THEC were exacerbated by the speed with which the schedule had been approved. The THEC staff had allowed little or no time for consultation with institutions or governing boards. Moreover, since instructional evaluation monies were not earmarked but embedded in the total appropriations request from the THEC, when that request was scaled back by the governor's office, institutions believed that they would be competing for dollars that might have come to them automatically from the regular funding formula. There were additional complaints that, first, the variables were not sufficiently responsive to two-year colleges or to the substantial graduate mission of the comprehensive universities and, second, that ambiguity and dysfunctions in the wording and construction of some of the standards made it difficult for institutions to respond to the schedule while simultaneously leaving members of the THEC staff open to charges of being arbitrary in the way they awarded points. These negative reactions swiftly led to the decision by the THEC to revise the schedule during the winter of 1979-1980. The revision process was an interactive one between the THEC, the institutions, and the staff of the two governing boards. After all of the objections to the then-current schedule were logged, the THEC approved a revised schedule in July 1980 (see appendix B). Some of the important changes are described below.

*Variable I [Accreditation].* The revised Schedule changed the previous range-of-percentiles method to a strict percentage allocation based on the number of accredited programs, that is, those eligible for accreditation by an agency of the Council on Postsecondary Accreditation (COPA). The governing boards agreed that the draft was a significant improvement, but registered some additional comments. First, there should be provision for the negotiation, explanation, or adjudication of accreditable programs that, for one reason or another, an institution might not wish to have accredited. (Some programs may appear—if only nominally—to be accreditable by a COPA agency, but in the institutional context they may not be accreditable without significant curricular and/or administrative alteration.) Second, some academic areas have no discipline-based accrediting organization, and it would be inadvisable to place any pressures on disciplines that are not now engaged in accreditation. Demands made by accrediting organizations already have had what some regard as an adverse impact on the campuses, since internal fiscal reallocations have had to be made to satisfy programmatic accreditation at the expense of nonaccreditable programs. Third, since instructional improvement is the aim of the schedule, the point values of each variable should be examined to see if a proportional weighting system could be agreed upon such that Variable I does not assume too heavy an importance in the overall schedule.

*Variable II [General Education Outcomes].* The revised Schedule reworked the various standards, tightened their language, and, for the first time, inserted a stipulation that institutions report curricular actions taken on the basis of the data gathered. The governing boards continued to note that defining and assessing general education were far more difficult activities than the THEC staff seemed to realize and that different institutions, particularly specific programs or colleges within a given institution, might have widely divergent ideas about what general education is or ought to be. Thus, in order to suggest the importance of the outcomes of general education and the difficulties of evaluating them, it was argued that heavier weight should be assigned to Variable II.

Two other comments of governing-board members bear mentioning. One is that the available instruments were norm-referenced tests designed to discriminate among individual test-takers and that, although the ACT COMP test appeared to be an appropriate criterion-referenced instrument designed specifically for the assessment of general education, its reliability and validity had not been fully established. Second, although improvement was clearly the goal of these evaluative activities, informed actions could not be taken on the basis of relatively short-term assessments, particularly in cases where assessment instruments were of questionable validity and reliability.

*Variable III (Program Performance Outcomes).* The revised schedule tightened the language of the standards, and added requirements involving the use of data that would have been very difficult to acquire, such as pretest/post-test data from comparable institutions. Members of governing boards commented that even in revised form this variable did not deal adequately with the problem of using standardized tests. Moreover, there was some feeling that comparative use of test data might have validity that was more apparent than real and that there was potential for misinterpretation and misuse of those data. While educators might recognize invidious comparisons among essentially unlike institutions, the lay public might not, and indeed, the schedule would be a very public document.

*Variable IV (Satisfaction Indices).* The revised Schedule reorganized and edited this variable without significantly altering it, and the governing boards applauded the result. The increased emphasis on recurring assessment was regarded as a positive step, since the pertinent literature suggested that only studies conducted repeatedly over time could be used as bases for informed changes. The use of community members/employers as a referent group was thought to be more appropriate for some institutions than others. Governing boards also suggested that nonreturning students, that is, "dropouts," could provide extremely useful data to an institution and that they should be included as an additional referent group.

*Variable V (Evaluation-Planning-Action for Renewal and Improvement).* The revised schedule refined the language of this variable without significantly altering its substance. Given the uncertainties the boards felt about Variables I, II, and III, it was suggested that special emphasis on campus planning be encouraged in Variable V. The governing boards advocated a planning-implementing-evaluating loop and urged that the variable be given a higher point value.

### Current Instructional Evaluation Schedule

By November 1981, nearly six years after the first steps were initiated in the Performance Funding Project, the Instructional Evaluation Schedule had been in use for two and one-half years, and clearly needed to be reworked. Different institutions continued to point out the same problem areas. These included unclear wording, lack of flexibility regarding institutional mission, and dysfunctions, including the weighting of points, between and among the variables' standards. Staff changes at the THEC, the governing boards, and the institutions, along with a reorganization of statewide governance of postsecondary vocational institutions, swiftly led to significant changes in the Instructional Evaluation Schedule. As a result of numerous lengthy meetings, the schedule was drafted and redrafted at

least four times between December 1981 and September 1982. In September 1982, it was approved by the THEC with the following notation: "Some temporary modifications to these revised standards may be necessary for the initial year of their application in order to facilitate the transmission from old to new evaluation concepts." A year was spent trying to reach agreement on those "temporary modifications," and even now some details of interpretation remain unanswered. The schedule approved in September 1982 so resisted interpretation, in fact, that it had to be reissued in "final form" once in June 1983 and again in November 1983 (see appendix C).

The current Instructional Evaluation Schedule is expected to remain in effect, with minor modifications as needed, for five years, by which time all programs will have been evaluated, a data base will have been built, and external reviewers will have studied the schedule's efficacy. Thus, while the conclusion for this chapter cannot be written until 1987-1988, it is now possible to provide a historical map of the thoroughfares, byways, and blind alleys of the schedule. A quick look at the appendixes makes it clear that much has already been learned. The current schedule now runs to nearly 12 single-spaced pages, whereas the original was only 2 double-spaced pages. And, while there are honest differences of opinion about the value of linking performance to resources, the schedule has generally been an effective agent of change in the state's institutions. Moreover, while representatives of campuses and the governing boards continue to be negative about some of the specifics in the schedule, the increase of its share of an institution's budget to 5 percent was accomplished by the THEC with little criticism in the fall of 1983. The design of each of the five variables in the current schedule can be summarized as follows:

*Variable I.* Program accreditation has been agreed upon as an acceptable statewide goal, but questions have been raised about the efficacy of some accreditors' standards. Tennessee institutions have been active participants in drafting nontraditional measures of outcomes for the regional accrediting association, the Southern Association of Colleges and Schools. Questions remain about the use of non-COPA agencies and about the advisability of forcing programs to fit accreditors' definitions. Is it appropriate, in other words, for an institution to change its policies or the curriculum of one of its programs in order to make a nominally accreditable program fit an accreditor's strictures? While it seems legitimate to use accreditation standards as one means of assessing the effectiveness of academic programs, assigning 25 percent of the schedule's values to this variable seems to be an excessive percentage, as well as an inappropriate application of accreditation standards.

*Variable II.* This variable requires that every eligible program be evaluated once during each five-year cycle of the schedule. There remain significant difficulties in identifying the external instruments that are appropriate to a specific academic program at a given institution. The difficulties involved in developing valid and reliable local instruments are

compounded because there is no way to apply results across institutions. More problematic still are the questions regarding external peer reviews. It is necessary to add information about student outcomes for consideration in peer reviews that are themselves already expensive, time-consuming, and cumbersome. Additionally, comparative assessments with either similar programs at other institutions or across time are very difficult and costly to obtain using peer evaluations. Finally, thoughtful analyses of most academic programs indicate that virtually all graduate, and many undergraduate, programs are tailored to the needs of each individual student. The THEC schedule does not permit an institution to attain maximum points via peer review, thereby discouraging the use of this potentially valuable assessment strategy. All of these difficulties notwithstanding, the presence of the Instructional Evaluation Schedule has focused faculty attention on intended program outcomes and discussion of the means for assesing those outcomes.

*Variable III.* It is fitting that the ACT COMP examination, field tested as part of the original Performance Funding Project, has emerged as the single instrument specified in general education outcomes assessments. There is, unfortunately, no additional reward for an institution's using the ACT COMP Composite Examination, which is expensive and difficult to administer but produces more reliable data for individuals and involves local faculty in scoring written and oral exercises. Variable III is also interesting in that it awards points based on gain scores as measured against either scores from comparable institutions or from previous administrations of the test. An alternative methodology suggested in the schedule—employment of graduates of nonbaccalaureate technical programs—is generally regarded as a less-than-satisfactory proxy measure. However, the presence of the variable has aided faculty discussions aimed at reaching consensus on the nature and purposes of general education. For example, the scarcity of explicit mathematical elements in the ACT COMP exam has made faculty interested in developing such items as part of the section of the exam that may be added by the institution. One can hope that the presence of a supposedly valid and reliable instrument will assist future development of curricula in general education.

*Variable IV.* The Performance Funding Project demonstrated that surveys of groups, including enrolled students, formerly enrolled students, or community members/employers, could, if used with discretion, be valid indicators of instructional effectiveness. This variable now carries a maximum of only 10 points. Its point value is small because questionnaires measure perceptions that, while real, are indirect indicators of instructional effectiveness. The variable still exists because the other assessments of student progress indicate only the degree to which students are capable of performing, not the level at which they perform, and because client assessments yield additional information to help explain scores on achievement tests. Experience seems to show that more attention should be given

to locally constructed surveys that separate responses according to level (perceptions regarding a program, department, college, or campus) and function (perceptions regarding classroom instruction, advisement, and academic supports such as library and computer resources).

*Variable V.* Perhaps the most important lessons learned from Tennessee's attempts to budget on the basis of instructional effectiveness have to do with the planning process. Being patient and persistent, having the broad-based involvement of institutional faculty and administrators, and linking assessment strategies to daily teaching/learning activities are all key elements. Thus, Variable V rewards institutions for having deliberate planning processes and assessing the extent to which these processes yield positive results. Programmatically, communication of the planning process should produce cooperation among institutions on local construction of tests and surveys, mutual assistance in identifying potentially applicable external instruments, and a sharing of strategies for administering the various assessment devices.

The THEC, governing boards, and Tennessee's institutions of higher education are still "acting on the possible while awaiting perfection," and even the current schedule relies upon indicators whose validity and reliability have yet to be established. Misunderstandings and points of contention still abound. The coda to this chapter cannot be written until the Instructional Evaluation Schedule, including its design, implementation, data base, and impact, has been thoroughly, carefully, and, impartially evaluated.

## References

Bogue, E. Grady; Harris, John; and Troutt, William E. *Allocation of State Funds on a Performance Criterion (Performance Funding Project).* Nashville, Tenn.: Tennessee Higher Education Commission (THEC), 1976.

Bogue, E. Grady, and Troutt, William E. *THEC Guide for Proposal Preparation.* Nashville, Tenn.: THEC, 1976.

Branscomb, Harvie; Milton, Ohmer; Richardson, John; and Spivey, Herman. *The Competent College Student. An Essay on the Objectives and Quality of Higher Education.* Nashville, Tenn.. Tennessee Higher Education Commission [THEC], 1977.

Fincher, Cameron. "Mid-Year Reports for Performance-Funding Projects." Tennessee Higher Education Commission, 1977.

Harcleroad, Fred. "Comments by Evaluation Team." Report to the Tennessee Higher Education Commission [THEC], 1976.

Levy, Robert A. "A Summary of THEC Performance Funding Project Directors' Meeting." Report to the THEC, 1976.

CHAPTER TWO

# Planning and Resource Allocation at the University of Tennessee, Knoxville: Institutional Responses to Changes in State Funding Policies

*By Homer S. Fisher*

DURING THE PERIOD of rapid growth in higher education, the emphasis of legislatures, as well as state and federal regulatory agencies, was much more upon accountability and control than upon the quality of higher education. Most public colleges and universities were subject to restrictive policies and procedures, rigid reporting requirements, line-item budgets, and careful scrutiny of positions and salaries during this era (Glenny and Schmidtlein 1983). Excessive attention of state legislatures to accountability of process, as opposed to concern for accomplishment, produced a myriad of regulations. Harold Enarson, then President of Ohio State University, captured the frustration shared by college and university presidents throughout the country:

> Legislators have rushed to regulate, I believe, in an attempt to restore the public's faith in political institutions. Their well-intentioned, if sometimes misguided, mandates are translated by bureaus and agencies into complex and often contradictory policies and procedures. Regulation, like the ash from Mount St. Helens, has filtered down from on high until it pervades almost every aspect of the university. And like the ash, it has choked life from some parts of the enterprise and made a mess of the rest. [1980, p. 9]

Enrollment-driven funding formulas have traditionally lagged behind increases in enrollment, they have failed to provide resources that adequately cover the increased costs of serving expanding student bodies, new demands for services, and extraordinary inflation rates. During the 1960s and early 1970s, most universities found themselves unable to match increases in growth with resources adequate to provide concomitant increases in quality (Hopkins and Massey 1981). During the 1970s, both economic limitations and changing societal priorities further detracted

from the funding base that supported higher education. During the period from the middle to late 1970s, enrollment stabilization or decline and reduction of in-state support resulted in inadequate planning, which in turn precipitated inappropriate across-the-board cuts, layoffs, deferred facility and equipment maintenance, and hastily developed financial exigency plans (Wittstruck 1982). Most universities have now recognized that these approaches do not retain or develop excellence, rather, they tend to foster mediocrity. Moreover, institutions are increasingly aware that the current period of enrollment (and price) stabilization, coupled with the renewed emphasis upon improving public education, should afford opportunities for addressing qualitative concerns as long as planning efforts and resource allocations are directed accordingly (Keller 1983) and state and federal regulatory agencies are willing to extend the necessary flexibility.

Within the last five years, state officials in Tennessee have started to emphasize resource allocation with the objective of improving the quality of programs and the performance of graduates. Some of the needed changes in focus have appeared gradually in modifications of funding formulas. The purpose of this chapter is to describe and evaluate the effects on the University of Tennessee, Knoxville (UTK), of changes in the state's funding policy for higher education.

## Traditional Approaches to Funding Formulas and Planning at Universities

State funding formulas generally employ cost and productivity (enrollment) factors to generate funding recommendations and/or to allocate funds among public institutions of higher education. Formula funding provides an objective, quantifiable approach to the allocation of funds that mitigates against institutional lobbying and reduces the pressures and tensions that result from utilizing political processes to secure resources. Although most funding formulas are, at least superficially, objective and quantifiable, they have major weaknesses. Cost factors by level and discipline are usually historically derived, thus they are frequently based on levels of program funding that, due to continuous shortfalls in appropriations, are inadequate at the outset. Statewide funding formulas often make little or no distinction with respect to role and scope among the kinds of institutions funded. Formula differentiation in these instances is based primarily on factors such as level of instruction and related enrollment, cost by discipline, and area of space to be maintained. This approach results in inappropriate recognition of the fact that different missions require different salary scales, unique equipment, special library collections, and other resources that are not considered adequately by some comprehensive statewide formulas. As Gross (1973) indicates, such efforts to provide equal treatment in funding all institutions of higher education can result in a leveling of quality.

College and university planning efforts traditionally have been undertaken at the departmental level or, if universitywide, through long-range, bottom-up planning that focuses upon internal rationales for growth, development of new programs, and allocation of resources (Keller 1983). During the past two decades, most colleges and universities have followed incremental budgeting processes. Occasionally, however, they have turned to planning, programming, and budgeting systems (PPBS), and/or the closely related concept of zero-based budgeting (ZBB). PPBS provides a framework for improving the relationship between planning and budgeting decisions, with emphasis upon setting program goals, objectives, and priorities prior to allocating or reallocating resources in accordance with these priorities (Parden 1972). ZBB is a form of program budgeting that requires, at least conceptually, that each existing program as well as each potential new program justify its continuation or initiation with each budget cycle. In effect, ZBB requires periodic reassessment of the priority of activities as well as the funding for these activities, including expanded resource allocations.

According to Schroeder (1973) and Hopkins and Massey (1981), no fully successful operational example of PPBS is available in higher education. The inordinate amount of detailed paperwork required by many such systems has been a barrier. Another limitation has been the analytical focus on the internal aspects of the institution to the neglect of effective examination of the external environment. Furthermore, the emphasis upon enormous amounts of quantitative data that assess such factors as program productivity as opposed to outcomes and other qualitative measures has also led to criticisms of PPBS. The PPBS concept, when expanded to include environmental assessment and the relationship of program decisions to the strategic decision areas discussed below, could have significant merit, particularly if program effectiveness is determined through a focus on quality that includes assessment of outcomes as well as processes.

### An Innovative Approach to Planning and Budgeting

An approach to budgeting based on strategic planning and program review provides alternatives to traditional methodologies. Examination of the external environment within which a university operates and assessment of internal strengths and weaknesses, with particular emphasis upon program evaluations and the setting of program priorities, are the two tenets upon which Shirley and Volkwein's (1978) model for strategic planning is based. According to this model, university officials should make, refine, or reaffirm decisions in six strategic decision areas, taking into consideration internal strengths/weaknesses and external opportunities/limitations, as well as the culture of the institution. The six strategic decision areas are the following:

1. The mission of the institution stated as specifically as possible. Priorities and limitations of the institution should be noted. The mission statement should provide realistic parameters for planning, without encompassing every conceivable activity that could be undertaken by the institution.
2. The goals and objectives that enable the university to carry out its mission.
3. The composition (size, quality, diversity, and level of study) of the student body that the university will serve, in accordance with its approved mission.
4. The range of programs and services to be offered and their relative size, quality, and funding priority.
5. The geographic service area in which the university will operate, both in terms of the campuswide mission and with respect to individual programs and services.
6. The comparative advantages that the university will seek to develop in its program areas.

Outcome information gathered through the process of instructional evaluation and used in program review is of particular significance in examining the composition of the student body and projecting changes, assessing program quality, and ascertaining the potential for comparative program advantages.

## The Case of Tennessee

Public institutions in Tennessee have not experienced the kinds of extreme regulation and control imposed by a number of other states during the past decade. Although Tennessee's colleges and universities are subject to appropriate fiscal reviews and audits, constraints such as line-item budgets, position controls, and the reversion of year-end funds have been consciously avoided by state officials. Furthermore, with relatively few exceptions, institutions in Tennessee may allocate and reallocate funds at the campus level. Funds may be shifted from category to category in accordance with institutional priorities. They can be shifted, for example, from operating budgets to salaries and vice-versa. Freedom from the restrictions of accountability requirements on one hand, and flexibility in resource management on the other, have not only reduced the administrative overhead burden for Tennessee's institutions but also contributed to a climate in which progressive changes in funding formulas and new ventures such as performance funding might be considered.

Both the current and previous Tennessee funding formulas have been more successful in developing funding requests than in allocating resources. This is primarily because resources have seldom been adequate to fully fund the formula recommendations at the levels generated. Accordingly, each

year for the past decade, funding recommendations for Tennessee institutions have been subject to some form of pro rata compression, with the result that most or all of the formula components have been underfunded for all institutions. Of course, no funding mechanism can meet the actual needs of an institution if the resources to be distributed are inadequate.

Underfunding has spawned criticisms, resistance to change of the formulas, and skepticism about innovations such as performance funding. Despite these concerns, however, in recent years positive changes have occurred in the Tennessee formula due to the incorporation of factors that improve the recognition of differences in the missions of state universities, four-year colleges, and community colleges. The changes include provision of supplemental funding for replacing equipment and special funding to improve endeavors regarding affirmative action, use of peer institutional reviews to establish funding goals, and—of greatest importance and impact upon institutional planning and program assessment—performance funding.

### How the Idea of Performance Funding Was Received at UTK

Although UTK and other state institutions were represented on the committee of the Tennessee Higher Education Commission (THEC) that developed the standards that govern performance funding, there has been widespread criticism of the common standards for all institutions, which inevitably are the product of numerous compromises worked out in a political environment. Academicians have raised questions about time and costs, as well as about the methodological and measurement issues involved (Kells 1983). Ostensibly the program was designed to promote improvements within institutions; in fact, it has sometimes fostered competition among institutions. Another apparent contradiction arises from the fact that while faculty members did not participate in development of the standards, they must be deeply involved in program evaluation.

The idea of performance funding was not initially greeted with enthusiasm by faculty or administrators in Tennessee. On the contrary, most members of the THEC committee for formula revision (a group of representatives from the two governing boards, their institutions, and the THEC staff) regarded the proposed new method of distributing or redistributing already limited funds as a serious threat that would further erode the inadequate funding base distributed by the conventional formula. However, after substantial discussion and a bit of informal moral suasion by some THEC members, the genuine desire of the committee members to continue to pursue means of introducing qualitative components to Tennessee funding decisions ultimately prevailed. Without question, mutual trust and confidence stemming from both the cooperative development of the basic Tennessee funding formula and the historical avoidance of inflexible fiscal constraints on the institutions were positive factors both in the decision to

adopt performance funding and in the subsequent implementation of the instructional evaluation program.

By 1983, acceptance of performance funding and instructional evaluation among institutions throughout the state was sufficient to permit virtually unanimous endorsement by institutional representatives of a proposal to increase the level of funds allocated through this mechanism from 2 percent to 5 percent of the instructional budget. Today, most institutions in Tennessee have welcomed the funds distributed through performance funding, even if they have not given complete endorsement to all of the instructional evaluation criteria employed. UTK has experienced significant financial benefit from the program, and the institutionalization of its components has already started to have an important impact on university planning. Utilization of outcome information from instructional evaluation activities in institutional planning and evaluation may have greater long-term significance for improved quality than will the budgetary ramifications.

## Implications of Performance Funding for Planning and Resource Allocation at UTK

The academic deans, faculty leadership, and campus administrators at UTK have started to utilize a model for strategic planning that embodies a number of the concepts discussed by Shirley and Volkwein (1978). Perhaps the most critical area of strategic decisionmaking involves the assessment of programs from the standpoint of their current and potential quality and their relationship to the external environment.

Annual budget hearings for departments and colleges at UTK provide the most frequent evaluations of program accomplishments, opportunities, and limitations. Participants in the budget hearings include the provost, (chief academic officer), executive vice-chancellor (chief fiscal officer), vice-provost and dean of the Graduate School, dean of research, and other central administrators. During the hearings, it is the responsibility of academic department heads and directors of major administrative/service units to outline the goals and objectives of their respective units, the relationship of those goals and objectives to the unit's and university's mission, the potential for achieving (or maintaining) excellence, and the associated resource requirements.

During these budget hearings, the information gained from the institution's evaluation of instructional activities contributes in significant ways to assessment of program quality and analyses of program priorities in determining resource allocations. Program characteristics examined in the course of the hearings include. (1) quality of faculty and support staff, (2) quality of facilities and equipment, (3) adequacy of library support, (4) assessment of student outcomes such as achievement in general education and in the major field, (5) assessments of program quality obtained from enrolled students, alumni, employers, and others, (6) relationship and

contributions of the program to the mission, goals, and objectives of UTK, (7) external demand for graduates of the program, (8) demand for courses by majors within the program; (9) demand for courses by nonmajors; (10) geographical advantages or disadvantages of the program; (11) comparative advantages that the program may have vis-a-vis programs at other institutions; (12) opportunities for sponsored research; and (13) cost-effectiveness.

Information obtained through these program evaluations is used to establish short-term, intermediate, and long-term budget priorities. Emphasis is placed upon the allocation of resources to strengthen less-than-adequate programs (where resource limitations provide barriers to accomplishing desired objectives) and allocation of increased resources to programs capable of achieving true distinction or in need of such resources to maintain distinction. The program evaluations are also essential in making decisions concerning long-term retention or elimination of programs, potential for program mergers, or reductions in program size or scope.

The program planning capabilities of UTK are further augmented by the campuswide assessment of student attitudes. The surveys of attitudes regarding adequacy of programs and services as reflected by the consumers of these services (see chapter 7) provide valuable information for central administrators concerning changes in perceptions over time, thus permitting improvements in services such as registration and advising or in-depth study of suggested problem areas. Furthermore, information about students' perceptions of the institution and their rationales for selecting UTK is beneficial to efforts to clarify the image of the institution and recruit students whose personal goals are compatible with those of the institution. Success in the latter endeavor, in particular, promotes retention.

Finally, the effectiveness of curricular planning is being improved as results of assessment of achievement and opinion are used to examine the degree to which various curricula are contributing to accomplishment of their stated goals. Recently, steps have been taken to include outcome information in program self studies and to improve the relationship between peer reviews and the annual departmental reviews conducted in connection with preparation of budget requests. Deans and department heads are asked to include information gained from prior peer reviews in the annual evaluations. Resource allocations are related to the results of both the relatively infrequent peer reviews (every five to seven years) and the annual departmental budget hearings.

## Conclusion and Recommendations

The new funding initiatives undertaken in Tennessee have attracted significant expressions of interest from representatives of institutions and governing boards in other states. While these inquiries have been stimulated in part by the publicity from the project sponsored by the National Center for Higher Education Management Systems (NCHEMS) and the

Kellogg Foundation, as well as from the recommendations of the National Institute of Education (NIE) Study Group on the Conditions of Excellence in American Higher Education, interest has also been spurred by a genuine enthusiasm on the part of many colleges for finding ways to institutionalize incentives for qualitative improvement. The potential for widespread utilization of outcome assessments at the statewide level is uncertain. The successful transition from traditional funding mechanisms to even the partial use of criteria related to performance in a meaningful way cannot occur overnight simply by decree of governing boards or agencies. The Tennessee experience suggests that several factors are facilitative, if not requisite, for statewide plans. (1) joint involvement of campus and governing-agency officials in initial planning; (2) significant participation by chief academic officers and faculty in developing methodologies, benchmarks, and reporting structures; and (3) institutional flexibility in designing specific plans for incorporating statewide criteria.

Instructional evaluation at UTK has moved from the conceptual stage to become an integral component of planning and budgeting. Individual institutions, both public and private, can benefit significantly from the use of outcome information regardless of whether it is mandated or supported by the state.

Campus resistance to innovative, qualitative funding initiatives will be reduced if faculty and staff can be assured that the new funds are truly incremental rather than merely a rearrangement of an existing funding base. If performance funding and the associated new funds had been introduced at a time when the formula-based request for higher education was fully funded, the new venture would have been more warmly applauded and provided improved opportunities to reward quality performance with new funds in a highly visible manner.

At the individual campus level, institutions should recognize the cost implications of the components of instructional evaluation and provide financial incentives to departments that participate in outcome assessments. Specific support should include funding of surveys, released faculty time for the development of testing instruments, logistical support for administering assessment activities, technical expertise as required, and preparation of reports that incorporate evaluation results clearly and concisely.

If the institution can develop a structure that assures recognition of program quality as demonstrated through outcome assessments, faculty, department heads, and deans will be much more willing to undertake the efforts required to institutionalize the use of this information. Such recognition can take several forms. Appropriate publicity and personal acknowledgment of a job well done by the institution's academic leadership will be far more beneficial than routine acceptance of outstanding performance. The development of continuing support by the faculty and departmental administration will be contingent upon the effective use of the outcome information in planning decisions that result in generally improved

approaches to the allocation of resources. Ideally, the ultimate reward should be the beneficial use of outcome information to achieve major goals of individual departments related to improved instruction, improved curricula, and better educated students. Indeed, program improvement must continue to be the highest priority goal of outcome assessment, annual departmental program evaluations, peer reviews, and related strategic planning.

Performance funding probably will never fully replace the productivity and cost considerations employed in funding formulas. Rather, performance funding, augmented by such categorical funding programs as centers of excellence, chairs of excellence, and programs for gifted students, will most likely provide additional means of improving quality for a number of institutions. The criteria employed will be subject to change as goals are achieved and as missions change, and definitions of quality will be elevated as institutions improve their programs.

## References

Enarson, Harold H. "Quality and Accountability. Are We Destroying What We Want to Preserve?" *Change* 12 (October 1980):7-10.

Glenny, Lyman A., and Schmidtlein, Frank A. "The Role of the State in the Governance of Higher Education." *Educational Evaluation and Policy Analysis* 5 (Summer 1983):133-53.

Gross, Francis M. "A Comparative Analysis of the Existing Funding Formulas Used for Justifying Budget Requests or Allocating Funds for the Operating Expenses of State-Supported Colleges and Universities." Ph.D. dissertation, University of Tennessee, Knoxville, 1973.

Hopkins, David S. P., and Massey, William F. *Planning Models for Colleges and Universities.* Stanford, Calif.: Stanford University Press, 1981.

Keller, George. *Academic Strategy—The Management Revolution in American Higher Education.* Baltimore, Md.. Johns Hopkins University Press, 1983.

Kells, H. R. *Self-Study Processes. A Guide for Postsecondary Institutions.* 2nd ed. New York: Macmillan, 1983.

Parden, Robert J. "Planning, Programming and Budgeting Systems." In *Efficient College Management,* pp. 10-19. Edited by William W. Jellema. San Francisco: Jossey-Bass, 1972.

Schroeder, Roger G. "A Survey of Management Science in University Operations." *Management Science* 19 (April 1973):895-906.

Shirley, Robert C., and Volkwein, J. Fredericks. "Establishing Academic Program Priorities." *Journal of Higher Education* 49 (September/October 1978):472-88.

Wittstruck, John R. "The Effect on Higher Education of State Actions in Response to Unanticipated Revenue Shortfalls. A Report of a Survey Initiated by the Washington Council on Postsecondary Education and Responded to by the State Higher Education Officers." State Higher Education Executive Officers and National Center for Education Statistics, Washington, D.C., April 1982.

CHAPTER THREE

# Comprehensive Program Evaluation at the University of Tennessee, Knoxville: A Response to Changes in State Funding Policy

*By Trudy W. Banta*

IN THE PICTURESQUE prose of George Keller (1983, p. viii), the image of colleges as "amiable, self-correcting collectives of scholars with a small contingent of dignified caretakers at the unavoidable business edge" is quickly crumbling. Program evaluation has become an imperative in higher education today, and it is increasingly being recognized as an important source of information for strategic planning and resource allocation.

Prior to the advent of state funding policies based on student performance, the University of Tennessee, Knoxville (UTK), like most other institutions of higher education, relied principally on the measurement of input variables to describe the quality of the educational experience for students. The financial carrot offered by the Tennessee Higher Education Commission (THEC) provided the incentive for faculty members and administrators to study outcome measures. UTK now has institutional policies and procedures for using information about student outcomes to evaluate and improve academic programs and make strategic decisions about program mix and the allocation of internal resources. Systematic collection of information about student outcomes, in other words, has been used to strengthen ongoing institutional activities.

The purpose of this chapter is to address the following question. In what ways has performance funding initiated by the state of Tennessee provided an incentive for comprehensive program evaluation at UTK? It is the assumption of this chapter that every institution of higher education can benefit from collecting evidence of program quality and using the results to make decisions and plan for improvements.

## Defining Comprehensive Program Evaluation

Evaluation can be defined either as a procedure for establishing worth of programs or as a procedure for providing information to be used in decisionmaking (Anderson, Ball, and Murphy 1975). Both definitions can be applied to higher education. Comprehensive program evaluation in colleges and universities can furnish evidence of program worth, thus satisfying external demands for accountability. Program evaluation can also contribute valuable information for internal decisionmaking. Comprehensive program evaluation monitors the setting of objectives, allocation and utilization of resources, and assessment of outcomes. It can answer questions such as the following: Which programs should be strengthened in order to achieve excellence? What are the essential services that need to be improved? Which programs should be phased out in order to provide the resources for strengthening and improving the programs that are more directly related to the mission of the institution?

Comprehensive program evaluation can be carried out in a number of different ways. George Kuh (1981) has developed a framework for evaluating the relevant variables. He differentiates context, input, involvement, and outcome variables. Context indices of quality are relatively stable characteristics of an institution's environment, such as size of student body, institutional mission, expenditures per student, and proportion of the faculty with a doctorate. Input indices are characteristics of entering students, such as intellectual ability, interests, aspirations, and demographic characteristics. Involvement indices include opportunities for students to interact with faculty and other students, student satisfaction with the institution, and the amount of time students spend on campus studying and becoming immersed in the educational experience. Outcome indices are the products or effects of the college experience, such as achievement levels (as measured by tests), cognitive and affective development, and postgraduate achievements.

## Methods for Measuring Performance Outcomes

Since no single evaluative instrument or process can provide all the information needed to assess the effectiveness of a given academic program, a combination of measures must be used. The National Institute of Education (NIE) Study Group (1984) has recommended surveys and comprehensive testing, and at UTK these assessment procedures have been combined with peer review to form a comprehensive process for evaluating academic programs. Tests provide answers to questions such as, Which content areas of the curriculum need to be strengthened? Questionnaire responses suggest how improvements can be made in processes that promote the learning of content, such as advising, classroom instruction, and grading. Peer reviews provide the means for considering the results of surveys and tests in combination with all of the other information about a

program, including quality of faculty, library collection, space and equipment. Peer reviews engender well-informed judgments about program quality.

The methods used to assess student outcomes must be right for the institution, that is, they must have been reviewed by administrators and faculty members and found to be content valid. In other words, they must measure those things the institution considers important. Additionally, the methods must be available at a cost the institution can afford.

When a decision is made to pursue academic program evaluation with an emphasis on performance, an institution should be prepared to provide strong administrative support and commit sufficient financial resources to mount a credible data-gathering effort. The goal of improving academic programs and services is broad; it should have an impact on every unit within the institution that serves students. Thus leadership must begin with the chief executive officer and that individual's staff.

The involvement of faculty members in the selection or development of assessment methods gives them a sense of ownership, which increases the likelihood that they will pay attention when responses are tallied and reported. Faculties that develop comprehensive exams should first agree on comprehensive behavioral objectives for majors. This may be very difficult in a department where heretofore the curriculum was simply that which resulted when each faculty member taught courses in his or her specialty, and the chief measure of comprehensive achievement was the cumulative grade-point average (GPA). After objectives have been ascertained, faculty members should develop test items to measure student achievement of the objectives. Faculty members should also take part in scoring or grading the resulting test. If they are involved in all of these preliminary steps, they will be interested in learning how students performed and initiate modifications in curriculum and instruction when test results indicate that these are needed to increase the quality of student achievement.

### UTK Study of the THEC Criteria

A Kellogg grant administered by the National Center for Higher Education Management Systems (NCHEMS) was awarded to UTK to fund a campuswide study of the THEC funding criteria with respect to student performance. Deans of the nine colleges that enroll undergraduates (Agriculture, Architecture, Business, Communications, Education, Engineering, Home Economics, Liberal Arts, and Nursing) were asked to name representatives to three task forces that spent six months carrying out extensive reviews of methods for measuring performance outcomes with respect to (1) student achievement in general education, (2) student achievement in the major field, and (3) student opinion concerning the quality of academic programs and services. In June of 1982, the task forces forwarded a set of recommendations to central administrators, which in the next 18 months were implemented in the following seven ways:

1. The American College Testing Program College Outcome Measures Project (ACT COMP) exam was judged to be the most appropriate instrument for evaluating studen progress in general education at UTK. A testing program was initiated by the Learning Research Center (LRC), a unit responding directly to the provost and charged with improving teaching and learning across the campus. The testing program involved administering the ACT COMP exam to representative samples of freshmen and seniors and comparing the two sets of scores to estimate value added by the college experience in general education.
2. Provost George Wheeler asked program heads to consider developing plans for administering to potential graduates a comprehensive exam in the major field at least once every five years. Program heads were encouraged to assess the validity of standardized instruments such as the Graduate Record Exam (GRE) Advanced Tests and professional licensing exams for evaluating student achievement in the relevant disciplines. Opportunities to secure the assistance of measurement consultants in the design of local tests were provided for departmental faculties that preferred to develop their own instruments.
3. Professors Kent Van Liere and William Lyons worked with central administrators, approximately 20 program heads, and a sample of students to design a questionnaire that would tap student opinion concerning, first, the adequacy of campus services, such as the library and the computer center, and second, the quality of the curriculum and instruction in the major field. The instrument initially was administered to a university sample of 2,200 enrolled students in May 1983. This survey provided data for the university, nine colleges, and five pilot departments. The questionnaire was used for the first time in January 1984 in the way that it was intended to be used on a continuing basis, that is, to sample student opinion in departments scheduled to undergo comprehensive program reviews during the current year.
4. In order to encourage departments and colleges to collect some of their own evaluative data and use it to make decisions about needed program improvements, a substantial portion of the Kellogg funds was used to finance pilot projects in several colleges. Each dean was encouraged to identify at least one project, and in January 1983, funds were awarded to 14 units representing all nine of the colleges enrolling undergraduates. Five departments used the GRE Advanced Tests to assess achievement in the major field, and two departments designed their own comprehensive exams. In order to determine student opinion concerning the quality of academic programs and services, four units pilot-tested the Student Satisfaction Survey developed by Van Liere and

Lyons, and three used surveys of their own design. Selection and administration of the tests and surveys, as well as review of the results, caused faculty members to view some aspects of their programs in a new light. In most cases the new insights stimulated interest in making improvements.

5. To ensure that representative samples of students would take the tests and complete the survey instruments identified for purposes of program evaluation, the assistance of deans and their representatives was enlisted to secure passage by the Undergraduate Council and the Faculty Senate of a requirement that every undergraduate participate in at least one evaluative activity prior to graduation. Options for fulfilling the requirement included the ACT COMP exam, a comprehensive test in the major field, and the Van Liere-Lyons survey instrument.
6. As a first step toward institutionalizing the use of outcome information at UTK, specifications for including outcome information in the self-study were added to the guidelines for the comprehensive program review. This type of action ensured that each academic unit preparing for a review would consider student achievement and assessments of the program by enrolled students, graduates, and/or employers as evidence of program effectiveness. This information would then be used in combination with traditional input measures, such as size of the library collection in the field and entering abilities of students, as the team of internal and external reviewers prepared a report on the strengths and weaknesses of the program following a visit of two and one-half days.
7. Information derived from program reviews was used by faculty, deans, and central administrators to make needed improvements. Members of the central administrative team that conducted the program reviews and related follow-up focused attention of faculty members and reviewers on the use of outcome information in the evaluation process. The UTK Planning and Budgeting Coordinating Committee, which is co-chaired by Provost Wheeler and Executive Vice-Chancellor for Business, Planning, and Finance Homer Fisher, started to use the documents pertaining to academic program review that emphasized information about student outcomes, in order to determine which programs should be terminated, which should be combined with others, and which should be strengthened through allocation of additional resources.

## Leadership: A Crucial Factor

Although performance funding is obviously a matter of interest to business officers in Tennessee, the author's survey of coordinators of instructional evaluation in the state's technical institutes, community

colleges, and universities revealed that, in all of the 17 institutions represented, the individual providing leadership for the response to performance funding is the chief academic officer or the president. The use of outcome information to improve academic programs and services can best be ensured if the enterprise is given strong support by the chief academic officer.

Instructional evaluation at UTK has been given high priority by the chancellor; the executive vice-chancellor for business, planning, and finance; the provost; and the vice-chancellor for student affairs. Table 1 provides an indication of the contributions to evaluation and instruction by the various officials responsible to each of the three members of the chancellor's staff.

Communication about the purposes, methods, and results of evaluation of academic programs is critical if faculty members and students are to cooperate in the evaluation process and successfully use the findings. Thus, as new testing and survey results become available, the LRC staff at UTK makes presentations to Chancellor Jack Reese, his staff, the Board of Deans, and the General Education Advisory Committee. At least once each year, the staff also meets with an appropriate faculty group in each college. This might include the entire faculty or a committee charged with improving curriculum and/or instruction. (The survey of the state's instructional evaluation coordinators indicates that many of Tennessee's other postsecondary institutions use these same channels to disseminate evaluation findings.) Additionally, the LRC staff holds luncheon meetings to inform support personnel in the deans' offices of the details of the instructional evaluation program so that these individuals are prepared to encourage faculty and student participation. Finally, the program is explained to students and their parents via orientation sessions and written communications from the Office of the Provost and the Office of Admissions and Records.

### ACT COMP and Other Exams

General education at UTK is not the responsibility of a particular department or program, rather, all units are concerned with its quality. Student achievement in general education is measured by the ACT COMP exam (see chapter 5). UTK is undertaking comprehensive statistical analyses to identify correlates of high achievement and high gain on the ACT COMP exam, and it is now testing every freshman and senior. The results warrant increased attention.

Information on ACT COMP exam scores, such as that included in appendix D, has been developed for the chancellor's staff, the Board of Deans, specific colleges, and the General Education Advisory Committee. Beginning in September 1983, the LRC staff prepared for each of nine colleges an annual summary of this outcome information. With the cooperation of deans, the staff scheduled a 30 to 45 minute session with the

TABLE 1

## Administrative Contributions to Instructional Evaluation at UTK

| Administrative Office | Contribution to Instructional Evaluation |
|---|---|
| *Provost* | |
| Learning Research Center | Overall coordination |
| Vice-provosts | Utilization of information in program reviews |
| Academic deans | Information utilization and dissemination to department heads and faculty |
| Faculty members | Information utilization and communication with students concerning instructional evaluation procedures |
| Secretarial staff | Communication with students |
| *Executive Vice-Chancellor for Business, Planning, and Finance* | |
| Planning & Budgeting Coordinating Committee | Information utilization in strategic planning and resource allocation |
| Office of Information Systems | Design of systems for handling student data |
| Office of Institutional Research | Data storage and retrieval |
| Office of Purchasing | Provision of payments to vendors of standardized exams |
| *Vice-Chancellor for Student Affairs* | |
| Dean of Admissions and Records | Enforcement of requirement that students participate in testing procedure |
| Records Office | Enforcement of senior testing requirement prior to graduation |
| Office of Student Data Analysis | Identification of students eligible for testing |
| Student Counseling Center | Administration of standardized tests and interpretation of scores for individuals |
| Career Planning and Placement Center | Administration of senior survey instrument |
| Orientation Office | Provision of information to entering freshmen concerning instructional evaluation procedures |

faculty of each college in order to present the findings and respond to questions. Presenting the results of outcome measures necessitated emphasizing that a single source of information should not be used by itself to suggest a precipitous change in curriculum or in the delivery of a support service. For example, it was pointed out that confirmation of weaknesses indicated by test scores should be sought in interviews with students or surveys of alumni opinion.

Faculty members associated with some 50 programs for which comprehensive standardized exams are available, including exams such as the Graduate Record Exam (GRE), National Teacher Exam, Engineer-in-Training Exam, and licensing exams in nursing, law, veterinary medicine, social work, and architecture, have started to look more carefully at student performance on tests in order to identify strengths and weaknesses of their own programs. In order to evaluate curricula, faculty in 50 additional programs have agreed to seek the help of content specialists outside the institution to develop their own comprehensive exams for majors. While these locally developed exams may lack the polish and some of the technical accuracy of standardized tests, faculty members who have committed their time and attention to development of these tests, as compared to faculty members who have chosen to administer a readily available test, are more interested in analyzing student performance and making needed improvements.

Program heads and faculty members at UTK have been asked to collect and use information about student achievement and student opinion in comprehensive program reviews that occur at intervals of five to seven years. At these times, ACT COMP exam scores and survey data for the unit are provided centrally, and the unit may supply test scores for potential graduates. The College of Business is one unit for which all of this information has become available simultaneously, and several changes have been implemented as a result of directions suggested by the data. High scores on the economics section of the GRE Advanced Test in Business, for example, convinced program planners that course requirements in economics for majors could be reduced slightly. This provided an opportunity to add a requirement in business law. Additionally, the survey of enrolled students revealed concerns about the availability of courses and limited contact with faculty. As a result, a comprehensive review of the college advising program was underaken, and faculty members were made more sensitive to the importance of personal interaction with students.

For three years, the total score of UTK seniors on the ACT COMP exam has been above the national average, and the value added (estimated mean score gain from freshman to senior year) by the experience in general education provided at UTK has exceeded the mean of value added that was calculated by ACT for a group of peer institutions. These findings obviously have not suggested the need for radical revision of the curriculum. However, the actual scores achieved by UTK seniors on two of the

subscales ("Functioning in Social Institutions" and "Solving Problems") are not as high as faculty representatives on the General Education Advisory Committee would like them to be. Consequently, Aubrey Forrest, director of the COMP at ACT, has been invited to the campus twice to present workshops designed to help faculty members increase the emphasis on higher order intellectual skills in all courses that they teach.

Interested faculty and administrators have developed a set of local items for addition to the ACT COMP exam that facilitate the identification of correlates of outstanding achievement in general education. Initial findings indicate that participation in student professional organizations is one such factor. This information has been shared with deans, and initiatives are under way in the agriculture, communications, and engineering colleges to increase participation in student organizations related to the major.

Use of scores on licensing exams and GRE Advanced Tests to suggest program improvements is limited in several fields by the absence of subscores in specialty areas of the discipline. Faculty members in the Department of Geography and the Department of Food Technology and Science worked for more than a year to design their own comprehensive exams for majors for the specific purpose of evaluating components of the curriculum. Both units benefited from the process of bringing together faculty members to discuss curriculum priorities and instructional objectives for students. Patterns of student strengths and weaknesses revealed by analysis of scores on the exams provided the impetus for faculty to modify some elements of the curriculum.

Some undergraduate units elected to design their own tests for administration and subsequent review in senior "capstone" courses for majors. Two departmental exams were developed as Kellogg pilot projects. Seven other units utilized funds from the NCHEMS-Kellogg grant to design their own survey instruments and administer them to collect data from students, alumni, and/or employers.

### Surveys

The work of one of the Kellogg task forces resulted in the identification of two UTK faculty members who were willing to involve students, program heads, and central administrators in the design of an opinion survey for enrolled students. Thus many individuals and units had a vested interest in the survey findings. Faculty members in the College of Engineering, the College of Business, the Department of Political Science, and the Department of Marketing and Transportation were sufficiently concerned about their students' reactions to undertake their own surveys to obtain more details about specific areas of dissatisfaction. One college added an advisor to the dean's staff because students indicated a need for better information about college offerings. The Department of Interior Design developed improved curriculum planning sheets for its majors on the strength of the assessment of advising by its students.

In designing the measures of student and alumni satisfaction with academic programs and services, Bill Lyons and Kent Van Liere interviewed students to determine which aspects of the academic experience they considered most important. After drafting instruments, they sent them out for review by selected faculty and department heads; they also asked that the section on university services be reviewed by the deans and directors reporting to the vice-chancellor for student affairs.

The faculty members and administrators who took part in developing these instruments have expressed interest in the survey findings, and they have used these findings to support initiatives for change. The results of a survey of student opinion, for example, were instrumental in effecting a number of changes in campus services. The registration process received relatively low ratings from students, and dissatisfaction with availability of courses during registration prompted the dean of Admissions and Records to institute an earlier deadline for admissions so that more information about completed schedules could be made available at the beginning of registration. Moreover, each dean was asked to provide a representative at the add/drop site who could counsel students at that point in registration.

The survey findings for specific colleges and departments have stimulated interest in making improvements in advising, internships, printed information about programs, and quality of instruction. Moreover, the results of testing and surveying have been incorporated into UTK's comprehensive program reviews, and the results of the reviews have been given increased attention and weight in strategic planning and resource allocation.

Data from the student survey were also used by Professors Van Liere and Lyons in a preliminary study of student persistence at UTK. Responses of students who completed survey instruments in May 1983 and then returned to UTK in September were compared with those of survey respondents who, for reasons other than graduation, did not return. Regression analyses revealed that overall satisfaction with experience at the university was the most powerful factor associated with persistence. Students with high GPAs and those employed for fewer than 30 hours per week were also more likely to return than those with low GPAs or those with 30 or more hours of employment. Patterns of student satisfaction were found to vary by college. The outcomes were summarized for the academic deans, and as a result, several of the deans planned appropriate responses.

### Impact of Instructional Evaluation

In perhaps the most significant action to date, the instructions used by program heads to prepare annual budget requests have been modified to suggest that information about student outcomes be included in the request as evidence of program quality. Additionally, the Planning and Budgeting Coordinating Committee has employed student outcomes, along with other

factors, as criteria to be used in selecting the campus proposals to be entered in Governor Lamar Alexander's statewide "Centers of Excellence" competition. Finally, the results of the student satisfaction surveys are scheduled to be used in the effort to increase enrollment and retention of black students as specified in a 1984 settlement of a lengthy desegregation suit that has implications for the entire state.

A survey carried out by the author in the fall of 1984 produced responses from 17 of 21 (81 percent) of the public institutions of higher education in Tennessee. The coordinator of the instructional evaluation program on each campus was asked to describe changes in academic programs or services that had taken place as a result of the instructional evaluation effort. Improved advising and increased emphasis on job placement were changes identified by administrators at all types of institutions. At four state technical institutes, additional changes included: (1) improved library services, (2) interaction and cooperation between faculty and the business community, and (3) increased faculty involvement in evaluation. Instructional evaluation coordinators at five community colleges noted the following actions: (1) course/curriculum changes in some majors, (2) increased faculty involvement in, and use of, evaluation, and (3) increased effort to attain accreditation of programs. At eight comprehensive universities, other changes were: (1) increased emphasis on use of comprehensive tests in evaluating programs, (2) new curriculum requirements for majors, (3) revised course requirements for majors, (4) extension of freshman orientation, and (5) termination of an unproductive program.

The coordinators were most frustrated with the part of the Instructional Evaluation Schedule related to comprehensive testing in the major field. They expressed concern about the costs involved, and they were troubled by the lack of appropriate norms for standardized exams. But they approved of the impetus that performance funding provides for undertaking evaluation, improvement, and renewal of academic programs, and for using evaluation results in planning and budgeting.

### Potential Problems

Even if the decision to pursue comprehensive program evaluation comes from within the institution, as it ideally does, arguments may be lodged against the approach that is taken. Faculty members and department heads typically feel somewhat threatened by program evaluation. Their fears may be expressed as objections to the proposed measurement methodologies. It should be granted that, indeed, no existing instrument or battery of instruments measures all of the intended outcomes of the college experience, including the abilities to reason, apply values, think critically, and solve problems. The instruments that are used most frequently—the GRE, the National Teacher Exam, and other specialized tests designed to certify competence for some agency external to the undergraduate institution—are pencil-and-paper tests that measure cognitive development almost

exclusively. Moreover, they are not content valid as criterion-referenced measures for use in program evaluation. These instruments are designed to measure the relative achievement levels of individual students, not to identify the specific strengths and weaknesses of a particular program at a given institution.

In some cases it may not even be possible to assess the content validity of an instrument because faculty members have not developed specific objectives for their programs. If the curriculum consists of a series of discrete courses the content of which is determined largely by the training and interests of the individuals who teach the courses, the faculty may not have agreed upon comprehensive performance objectives for majors. The cumulative GPA is considered the best indicator of student performance, and the absence of measurable comprehensive objectives makes it impossible to select or develop an examination with content that is valid for evaluating the curriculum.

Most faculty members have not had training in measurement theory or experience in developing standardized tests (Finn 1984), and this makes it unlikely that they will undertake development of a local test. Moreover, test development is a time-consuming task even for measurement specialists; there is little motivation for faculty members to spend time in this way because promotion and tenure decisions are not influenced as heavily by quality of teaching or attempts to improve instructional techniques as they are by quality and quantity of scholarly publications.

Suggesting that program quality be evaluated through testing always provokes the speculation that instructors will begin to concentrate their teaching on the content of the information covered in the test. It should be noted that this may not be a negative result if the measurement procedure is comprehensive to the extent that it assesses most of the student outcomes agreed upon as being important by the program faculty. Finally, academicians will point out that resources for higher education are limited and that the time and money devoted to selection of a test or to test development and administration could be used instead to make immediate improvements in the instructional program.

### Lessons Learned from the UTK Case

The study of performance funding at UTK resulted in a number of recommendations for using the instructional evaluation activities of the THEC to strengthen ongoing institutional activities such as program evaluation, improvement of curriculum and instruction, strategic planning, and allocation of internal resources. The implementation of these recommendations has involved all members of the chancellor's staff, as well as all academic deans and department heads. Ultimately, every student and faculty member at the university will become involved as a participant in data collection and/or as a recipient of the information derived from such activities.

The generally favorable response to the instructional evaluation program at UTK has been due in large part to the positive approach used by central administrators in examining its implications carefully with input from many faculty members and in making appropriate suggestions about implementation for the UTK campus. Department heads and faculty generally have responded to the new initiatives with openness, receptivity, and the expectation of constructive change.

Not everyone has reacted favorably to performance funding; some have been heard to grumble, "Just more busywork from the THEC." Well-grounded objections to, and legitimate concerns about, the instructional evaluation standards have been voiced and are explored in subsequent chapters of this book. The negative sentiment has not prevailed, however. The campus ethos incorporates a strong commitment to effective instruction of students. It is recognized that at the heart of the THEC program there is a promise of improved teaching and learning.

As the assessment program matures, more emphasis can be placed upon evidence that quality of outcomes has improved, but not simply in terms of test scores. If, as the NIE Study Group (1984) recommends, each program defines its own comprehensive objectives and standards of performance for majors and assesses student achievement accordingly, the need to improve components of the curriculum and instruction will emerge. Suggestions for improving the delivery of instruction, advising, and other academic services related to the major can be derived from surveys of students, former students, internship supervisors, and employers. In a peer review of this information, a set of specific recommendations for program improvement can be developed, and progress toward implementation of these objectives can be monitored. The program faculty can establish its own criteria for evaluating the effectiveness of improvement initiatives, and performance funding can be based on the extent to which these criteria are met.

At UTK, to summarize, performance funding has provided the incentive for incorporating outcome information of several kinds in comprehensive program evaluation that has acquired new status in the internal processes of planning and resource allocation. This experience may be instructive as other institutions, motivated by important factors influencing higher education today, undertake the complex process of building their own evaluation systems based on performance outcomes.

## References

Anderson, Scarvia B., Ball, Samuel, and Murphy, Richard T. *Encyclopedia of Educational Evaluation.* San Francisco: Jossey-Bass, 1975.

Finn, Chester. "Trying Higher Education. An Eight Count Indictment." *Change* 16 (May/June 1984):29-33.

Keller, George. *Academic Strategy. The Management Revolution in American Higher Education.* Baltimore, Md.: Johns Hopkins University Press, 1983.

Kuh, George D. *Indices of Quality in the Undergraduate Experience.* AAHE-ERIC/Higher Education Research Report, no. 4. Washington, D.C.: American Association for Higher Education, 1981.

National Institute of Education [NIE]. *Involvement in Learning. Realizing the Potential of American Higher Education.* Report of the Study Group on the Conditions of Excellence in American Higher Education. Washington, D.C.: NIE, 1984.

PART TWO

# Measurement Issues in Performance Funding

CHAPTER FOUR

# Accreditation as a Performance Indicator

*By C. Warren Neel*

PERFORMANCE FUNDING is part of an increasingly widespread effort to hold educational institutions accountable to the publics they serve. Since the 1960s, an increasing number of state and federal programs have been employing accountability measures to determine how funds should be allocated among competing programs. Performance funding, as developed by the Tennessee Higher Education Commission (THEC), uses performance indicators not only to make allocation decisions but also to encourage improvement of educational quality. Standards governing performance funding include program accreditation because it is recognized as an assessment of the quality of education by peers within the academic community. Indeed, as one of the oldest established assessment mechanisms in higher education, accreditation represents a particularly appropriate foundation for other models. Yet problems arise when institutions attempt to use accreditation as a performance indicator.

A paradox exists between what is now being asked of education in the name of accountability (outcome measures) and traditional accreditation standards (input measures). To date, little effort has been made to incorporate measurement of outcomes in program accreditation procedures. The emphasis, rather, is on input characteristics of minimum standards for admission, curricular requirements, and faculty credentials. This chapter outlines the history and purposes of accreditation, and identifies the issues connected with its use in performance funding.

## History of Accreditation

The first American institutions of higher education were founded primarily to serve relatively narrow interests such as the education of the Protestant ministry. They supported classical curricula for the education of an elite few. During the mid-1800s, there arose a growing demand that

colleges meet the needs of a broader spectrum of the population in order to justify increasing levels of federal and state financial assistance. By 1862, the pressure had become so great that the Morrill Act was passed. The Morrill Act mandated that in each state at least one university be responsible for teaching agricultural and mechanical arts. The Morrill Act thus strengthened the emerging philosophy that higher education is for every qualified individual. The classical curriculum was breeched, and programs of study directly related to the needs of constituents became the driving force behind a new standard curriculum that served "the people."

During the last 25 years of the 19th century, as institutions of higher education began to compete for students, many colleges obtained charters and recognition from governmental bodies in order to attract students. Regional variations in curriculum became the norm as institutions attempted to differentiate their programs from those of competitors. There was so much diversity among curricula that high schools were hard pressed to prepare students for the different admission exams required by different universities.

By the last decade of the 19th century, it was clear that some kind of standardization was needed. Higher education had become comfortable with the laissez-faire environment. Institutions were isolated, and identifiable standards that reflected the uniqueness of one institution did not—perhaps for reasons of geography alone—impinge on the standards promulgated by another. But as the country shifted from a rural, farm economy to an industrialized, urban economy, differences became more pronounced and conflicts arose. The social costs of maintaining incongruent standards became too great to tolerate. C. M. McConn, then dean at Lehigh University, summarized the prevailing mood of the era in the following way: "American education ... had come to be a variegated hodgepodge of uncoordinated practices ... which had never undergone any screening from anybody, and many of which were shoddy, futile, and absurd beyond anything we now conceive of, and the Age of Standards brought order out of chaos" (Selden 1960, p. 28). From this chaos came the first regional, then national, organizations that assumed responsibility for accrediting institutions of education. Since the formation of regional associations, there have been many transitions in process and substance. The premises of accreditation have not changed, but the responses of accreditation to a changing society have become apparent (Young, Chambers, and Kells 1983).

In recent years, public pressure exerted through the courts and other channels has led to an increasing emphasis on the public accountability aspect of the accreditation process (Harcleroad and Dickey 1975). This focus is a manifestation of the consumer movement that has led governments to demand clear evidence of the worth of product and service outcomes of social programs, including education. Government agencies have come to expect accreditation standards that reflect a direct correspondence between social investment and return and that serve as a means

of certifying minimum quality standards for the consumer (Orlans 1975). The Study Group on the Conditions of Excellence in American Higher Education of the National Institute of Education (1984) has recommended that accrediting agencies require evidence of not only the number of degrees awarded by an institution but also the proficiency gains of its students.

Several regional accreditation agencies have, in fact, acted on this advice and have started to encourage institutions to include information on student performance in accreditation self-study reports. The Southern Association of Colleges and Schools (SACS), for example, recently adopted a standard for institutional peformance that requires that each college or university provide a statement of its objectives and a body of evidence that these objectives are in fact being achieved. Other regional accrediting agencies—most notably the North Central Association and Middle States—are encouraging institutions to include information on student learning and development in their self-studies (Thrash 1984).

Paralleling the history of institutional accreditation has been the rise of specialized professional accrediting associations. The purpose of such associations is to certify the quality of professional training programs by ensuring that the curricula of such programs conform to established standards or that the performance of program graduates is of sufficiently high caliber to ensure effective participation in the profession. The most recent inventory of specialized accreditation bodies compiled by the Council on Postsecondary Accreditation (COPA) lists over 50 fields subject to accreditation of this kind.

Specialized accreditation bodies have traditionally devoted greater attention to program outcomes than have the regional bodies that accredit institutions. Different specialized accrediting bodies have different standards, of course, and they assign various degrees of importance to the factors considered in the accreditation process. However, all self-studies and subsequent external reviews are expected to be thorough assessments of the current state of the program as an aid to quality control and to the further development of the program. Recent trends in specialized accreditation have been toward greater standardization of data and greater emphasis on actual student performance (Christal and Jones 1985). Some standards, for example, require that students demonstrate their ability as a condition of graduation—often through a standardized examination or through the performance of acquired skills.

Accreditation has served the education community in this country in a variety of ways during the past 100 years. Its purposes remain laudable. The protection of program diversity while establishing minimum standards, coupled with the thrust to encourage institutional improvement, will remain goals of the scholar and of those individuals responsible for allocating taxpayer funds to nurture academic excellence.

## Purposes of Accreditation

The current accrediting process has four fundamental objectives that have remained steadfast as the underlying rationale for accreditation (Millis 1957): (1) maintenance of standards for admission to the institution, (2) maintenance of minimum academic standards, (3) stimulation of self-improvement, and (4) assurance that no countervailing forces are present at the institution. The process involves auditing the evidence (as presented in a self-study document) that an institution is following the standards it has promised its various publics. Regarding specialized or programmatic accreditation objectives, reviewers look for evidence that "students are exposed to certain learning experiences, professional practices and common bodies of knowledge presumed to be necessary for professional practice" (Young and Chambers 1980, p. 96).

Evaluation of the admission function focuses on the institution's public stance. The accrediting agency looks for clear statements reflecting an admission policy and for evidence that demonstrates adherence to the policy. For the purpose of evaluating the institution's maintenance of minimum academic standards, accrediting bodies typically compare catalog requirements and student records to determine the degree of congruence between stated admissions criteria and those actually enforced. The impetus for maintaining minimum academic standards derives from efforts by accrediting bodies to bring order and uniformity to students' educational experiences while protecting program diversity. In order to balance these conflicting goals, accrediting agencies typically focus on minimum standards, thus avoiding encroachment on program diversity while encompassing the majority of potentially accreditable programs.

The process of internal review and monitoring of minimum standards and program resources is designed to stimulate each institution to higher levels of excellence. Peer reviewers combine the results of their assessments with the results of internal studies to enable institutions to improve beyond the minimum level. They also attempt to ascertain if the institution has countervailing forces that impinge upon its mission and independence in determining and administering standards of academic performance.

## The Use of Accreditation as a Performance Criterion

In the performance standards specified by the THEC, accreditation is used, along with several other criteria, as a measure of institutional quality. While the THEC model offers significant improvement over other subjective methods historically employed as bases for justifying program allocations, there are problems associated with the use of accreditation as a surrogate measure of program quality. In the THEC program, an institution with 100 accreditable programs, 75 of which are accredited, receives 75 percent of the allocation for accreditation in the Instructional Evaluation Schedule

that forms the basis for performance funding. A larger percentage of accredited programs suggests a better quality institution. This is the basis for the contribution of accreditation to the Tennessee model and the focus of the following analysis.

One purpose of accreditation, as noted previously, is maintenance of admission and academic standards. Because most of these program features constitute input variables, in many cases they do not make a statement about the quality of program outcomes. An institution with an open admissions policy and admission standards that reflect this policy will meet relevant accreditation standards. A second institution that is highly selective in its admissions standards and has a student body that reflects that selectivity will also meet relevant accreditation standards. Each mission for each institution is accommodated in the accrediting process, but the two institutions, with their very different student populations, may vary considerably in academic quality as indicated by outcome measures.

An institution with more selective admission standards, for example, may choose not to seek accreditation for one program because it means duplicating resources already present in other accredited programs. An institution with an open admissions policy might, on the other hand, choose to accredit all its programs, spreading resources thin and duplicating departments, courses, and faculty in order to maximize its claim to state funds through performance funding. The less selective institution would receive more credit from the state for accreditation, even though its pursuit of accreditation for all programs could jeopardize overall institutional quality.

One common requirement for accreditation of a professional program is that a given percentage of full-time faculty in a college teach the courses offered for a degree in that college. For example, if the business school on campus is accredited, and if a hotel and restaurant management program in a home economics college on the same campus is also accredited, each program has achieved this designation in part because faculty members from the unit offering the degree teach the courses that make up the program. Thus, in order to meet two sets of program-accreditation standards, similar courses must be provided from separate budgets. Principles of economics and management taught by faculty in each school may be identical (even the textbooks and materials may be the same), but the accreditation process demands that the faculty teaching each program be governed by the dean and the faculty of the college granting the degree. For this reason, pursuit of accreditation may be a mixed blessing for an institution seeking to maintain an integrated academic program. Furthermore, some curricular requirements established as responses to accreditation standards virtually crowd out of a student's program important courses in general education. If an institution wishes to maintain a core in general education, the choice is often between refusing to seek accreditation and adding a fifth year to a student's undergraduate program.

It may well be that in the best institutions accreditation will not be pursued for every program that is accreditable, because to do so would not serve the interests of establishing quality throughout the institution. The highest quality institutions husband their resources, moving them from one area to another as the need arises without regard for the artificial constraints of accrediting agencies.

A third purpose of accreditation is institutional self-improvement. A college or university may attempt to improve its programs by changing standards and expectations. Such action is facilitated by ample resources and the flexibility to apply them to places of critical need. If undue credit is placed on accreditation in performance funding, presumably every program will pursue accreditation. Then funds will be expended to meet minimal accreditation standards in every area, and little or nothing will be left for effecting improvements. If this happens, a major thrust of the accreditation process is lost.

## Conclusion and Recommendations

Performance funding is a useful concept, yet it cannot rely too heavily on accreditation as a surrogate for quality. The accreditation process should not be called upon to do more than it was originally intended to do. If its historical objectives and current reasons for existence are ignored, it becomes a tool of external management.

There is little doubt that performance funding should include accreditation, but it should go further to achieve the purpose of encouraging institutional self-improvement. One suggestion is to incorporate into the model program reviews that focus on assessment of student outcomes that are carried out by respected scholars from outside the institution. As program reviews are presently used at the University of Tennessee, Knoxville (UTK), they are relied upon by department heads and academic deans to provide a basis for strategic action in curricular design and funding reallocation. Because the reviews are action oriented, target goals are set and specific improvements are articulated. Follow-up meetings are held within one year of the consultants' review and include documentation of changes that have been made, improvements that have resulted, and suggestions that have not been implemented. Administrators and faculty members of accredited programs and of those programs that are not pursuing accreditation, as well as those programs not eligible for accreditation, gain valuable insight for improvement through such nontraditional reviews. Comprehensive program review can therefore improve the functioning of performance funding by differentiating programs above the minimum acceptable standards set by accrediting agencies.

Despite its limitations, accreditation has some face validity, or surface credibility, as a criterion for program evaluation that determines the allocation of a portion of state resources. Accreditation is simple and straightforward for an institution to report, and it is easy for the state coordinating

agency to interpret. As an indicator of at least a minimum level of quality, especially if measures of student outcomes are included, the inclusion of accreditation in standards for performance funding is acceptable. Based upon research findings at UTK and the comments of administrators and faculty members at other institutions throughout the state, however, it is suggested that there be far less weight assigned this factor in the THEC Instructional Evaluation Schedule. Criteria for performance funding may therefore include accreditation, however, the following caveats should be kept in mind:

1. Accreditation generally focuses on input, not output, measures. Accreditation of a program means that minimum standards have been achieved; beyond that minimum, the process does not differentiate program quality.
2. Program reviews often provide the basis for significant program improvement, while the traditional accreditation self-studies and reports are too general or do not focus on improvements beyond a minimum acceptable level.
3. An institution's mission may suggest that some of its academic programs should not seek accreditation because resources allocated to that effort would detract from other programs that have already achieved a level of quality beyond the minimum certified by accreditation.
4. When accreditation is given too much weight in a performance model, it may force an institution to make choices for funding that bring program quality to the lowest common denominator rather than encourage diversity and excellence.

Accreditation, to summarize, should be considered a useful but inherently flawed criterion. A process that merely establishes that a program has met certain minimum standards cannot be a very powerful indicator of program quality. Moreover, at any given institution there may be effective programs for which faculties and administrators elect not to seek accreditation. Each institution has its own unique mission and program mix based on the needs of those it seeks to serve. A program viewing itself as part of a broader unit that tries to provide students with a rich experience in general education may choose not to pursue accreditation from an agency that would require its students to spend a much larger proportion of time acquiring professional competence through courses in the major. An administration faced with the prospect of having to duplicate core faculty to gain accreditation for each of two closely related programs housed in different units may exercise prudent judgment in electing to accredit only one of the programs. Finally, a program may be designed to meet certain job-related needs of a population that the institution wants to serve, and changing the program to meet all of the requirements for accreditation could render it dysfunctional for its local clientele.

## References

Christal, Melodie E., and Jones, Dennis P. *A Common Language for Postsecondary Accreditation. Categories and Definitions for Data Collection.* Boulder, Colo.: National Center for Higher Education Management Systems [NCHEMS], and the Council on Postsecondary Accreditation [COPA], 1985.

Harcleroad, Fred F., and Dickey, Frank G. *Educational Auditing and Voluntary Institutional Accrediting.* Washington, D.C.: American Association for Higher Education, 1975.

Millis, John S. "Major Purposes of Accrediting." In *Report of Workshop Conference on Accreditation,* n.p. Washington, D.C.. National Commission on Accreditation, 1957.

National Institute of Education [NIE]. *Involvement in Learning. Realizing the Potential of American Higher Education.* Report of the Study Group on the Conditions of Excellence in American Higher Education. Washington, D.C.: Government Printing Office, 1984.

Orlans, Harold. *Private Accreditation and Public Eligibility.* Lexington, Mass.. D. C. Heath and Company, 1975.

Selden, William K. *Accreditation.* New York. Harper and Brothers, 1960.

Thrash, Patricia R. "Accreditation and the Evaluation of Educational Outcomes: A Regional Accreditor's Perspective." Paper presented at a professional development session sponsored by the Council of Specialized Accrediting Agencies [CSAA] and COPA, Chicago, April 1984.

Young, Kenneth E., and Chambers, Charles M. "Accrediting Agency Approaches to Academic Program Evaluation." In *Academic Program Evaluation,* pp. 89-103. New Directions for Institutional Research, no. 27. Edited by Eugene C. Craven. San Francisco: Jossey-Bass, 1980.

Young, Kenneth E.; Chambers, Charles M., and Kells, H. R. *Understanding Accreditation.* San Francisco: Jossey-Bass, 1983.

CHAPTER FIVE

# Measuring Achievement in General Education

*By W. Lee Humphreys*

INCLUSION OF achievement in general education as one of the variables to be assessed in the Instructional Evaluation Schedule of the Tennessee Higher Education Commission (THEC) provided an external and very real stimulus for the University of Tennessee, Knoxville (UTK), to take advantage of current nationwide trends in the renewal of general education in baccalaureate programs and move beyond the experiences of most institutions in assessment of student achievement in general education. With aid from the National Center for Higher Education Management Systems (NCHEMS) and the Kellogg Foundation, an investigation was undertaken to discover ways to measure student attainment with respect to basic goals established by a UTK coordinating committee for general education. The purpose of this chapter is to review some current efforts to improve general education, including the use of information about student outcomes.

## Renewed Interest in General Education

In 1977, the Carnegie Foundation for the Advancement of Teaching declared general education in this nation's colleges and universities to be a "disaster area" (p. 164). At that time, few academics stepped forward to disagree with the assessment. Indeed, few were able to define with any specificity the goals of general education contained in the curricula of their institutions. This is an area in which judgments have often been based on hunches and recollections of one's own college days, and indeed, discussions of general education have traditionally been anecdotal. In the brief time since the Carnegie pronouncement, however, efforts to correct the situation have been launched. On the national level, these efforts are contained in various commission reports, studies, monographs, articles, proposals, and workshops, as well as in new avenues of funding. Perhaps

less visible but more striking are the efforts of many colleges and universities to assess the state of general education and, in many cases, to chart new directions in their curricula.

For the past 100 years, higher education has increasingly been organized into discrete disciplines, these, in turn, are organized according to departments, majors, graduate degrees, and specialized research. General education, in contrast to the clear organization of disciplines, is typically treated as everyone's concern and, too often, no one's responsibility. The geographical layout of most campuses reflects this, general education has no specific turf, office, department, college, budget, or administration. No one speaks primarily for general education. It has no substantive place in the organizational chart of an institution and no line in anyone's budget, and too often in the eyes of faculty members, it does not offer the usual rewards. Most important, the content of general education is not addressed at the initial and most formative level of course design and program development, that is, at the department level.

Changes are emerging, however. The awareness of the historical shift toward rarefied disciplinary divisions, together with its impact on the structures and processes of the university (decreased mobility for faculty and renewed demands for accountability), have resulted in more attention for general education. With respect to the individual faculty member, it no longer makes professional sense to define oneself solely by the criteria of a discipline. It no longer makes sense to treat an academic institution solely as the place where one serves a discipline until its demands necessitate that one move to another academic institution. Presidents, deans, and other academic leaders have been increasingly willing and able to assert their own values and visions of what should constitute an academic program.

A survey conducted in 1981 by the Association of American Colleges (Gaff 1983, pp. 197-206) indicates that the following are among the highlights of the current wave of renewed interest in general education.

1 General education in the curricula of both liberal arts and professional programs has increased at the expense of electives. A more careful definition of the goals of general education has led to a tightening of requirements for the broad distribution areas characteristic of the 1960s and early 1970s.
2 Courses designed to meet requirements of general education typically must receive approval of campuswide or collegewide committees and be assessed by criteria that are not specific to a single discipline. Thus, while few institutions have established a core curriculum in general education for all students, some institutions have proposed limited cores or they have emerged in larger distribution schemes.
3. Writing skills, problem analysis, problem solving, and other dimensions of critical thinking have received particular attention

in the articulation of goals; and foreign languages, the arts, and the humanities have regained some ground lost in the past several decades.

4. Courses are being designed to introduce students to higher education and facilitate academic planning and acquisition of basic learning skills. General education is also being extended into the third and fourth years with new attempts to articulate the relationship between liberal education and the major or professional program.

In his overview of the status of general education in the nation's colleges and universities, Gaff (1983) lists about 300 institutions, including public and private institutions, community colleges, graduate and research institutions, that are engaged in some type of review and reform of general education. Gaff's list makes it clear that, in just over half a decade, the most recent wave of renewal in general education has neared its crest.

Gaff suggests that reform of general education in institutions across the country is at different stages of development. (1) a few remain unaware of the ferment surrounding the topic, (2) many are reviewing their programs, usually through a task force established for this purpose, (3) a number are implementing new programs or major revisions of existing programs, and (4) a few are seeking to maintain or fine-tune new programs. Most institutions are in the second and third stages, relatively few have advanced to the fourth. This means that evaluation of student achievement that results from the most recent efforts to reform general education is itself in a nascent stage. Evaluation should play a role in all stages of the reform process, but it is especially important in the assessment of revisions and new programs.

### Assessment of Achievement in General Education at UTK

The third variable in the THEC Instructional Evaluation Schedule centers on the assessment of student achievement in general education. As applied to UTK, this variable states:

> (1) The institution will be awarded 5 points if within the past five academic years the institution has assessed the performance of a representative sample of graduates for its major academic degree utilizing the ACT COMP [American College Testing Program College Outcome Measures Project] Objective or Composite measure.
> (2) The institution will be awarded 20 points if through annual assessment utilizing the ACT COMP measure, the institution can demonstrate achievement of above-average performance in value added in general education outcomes when compared to a similar set of institutions. Or, the institution will be awarded 20 points if through annual assessment utilizing the ACT COMP measure, the

> institution can demonstrate an improvement in value-added from the most recent institutional measure of value-added. [Tennessee Higher Education Commission 1983]

Coincident with the state's development of performance funding, UTK was engaged in an internal review of general education. In 1979, in the annual address to the faculty, Chancellor Jack E. Reese called for an assessment of general education at UTK. A coordinating committee on general education, composed of faculty members from all colleges granting baccalaureate degrees, was formed to assess the current state of general education and to formulate the goals that should inform all baccalaureate programs. At the same time, ad hoc committees were established in each college to act as liaison between the coordinating committee and the faculties of departments and colleges. The process of review and definition by the coordinating committee made apparent the historical, structural, and procedural barriers to the implementation of new or revised general-education programs in institutions such as UTK.

The coordinating committee developed a statement of standards made up of three components: (1) basic skills, (2) areas of knowledge and patterns of inquiry, and (3) attitudes and perceptions. The coordinating committee mandated that all undergraduate curricula at UTK should strive to develop these components and that all should be assessed by these standards. The ideals were to be understood not simply as a base upon which a professional concentration or a major would be built (a perception of general education that faculty members who are oriented toward a discipline too easily embrace), but also as valuable in their own right because they prepare students for creative, rewarding lives and responsible participation as citizens of the nation and the world. The committee recognized that the ideal standards could be attained through a variety of courses drawn from both liberal arts and professional colleges. The committee also stressed that the line between liberal learning and professional education should not be too sharply drawn. The latter should build upon the former and reinforce the qualities that define it.

The UTK coordinating committee's statement of ideals attained broad consensus across the campus and was recognized for what it was—a statement of ideals. Actual curricula are designed in less-than-ideal settings. Typically they must operate under constraints, including those imposed by four-year baccalaureate programs, the remedial needs of some students, state or national standards for certification, and program accreditation. While such statements typically fall short in some way, this one provided a clear guide for the evaluation and revision of courses and curricula. Questions remained nonetheless. How successfully were the goals being attained? What difference did the reform and the new programs make?

## Use of the ACT COMP Exam in General Education

A task force was created to review available instruments that purported to measure student attainment in general education. The group selected the ACT COMP exam as the measurement instrument best suited to current UTK needs and resources. Critical factors that shaped the decision to use the ACT COMP exam included the extent of its coverage, its cost, and the time involved in scoring and interpretation, as well as the fact that programs and not individual students were being assessed. Materials available from McBer and Company (Winter, McClelland, and Stewart 1981) were considered because they had genuine strengths and incorporated areas not treated by the ACT COMP exam. They were ultimately rejected, however, due to the time and cost involved in scoring them and because they were designed more for the assessment of individuals than programs. The task force also considered using the Graduate Record Exam (GRE), but they found that verbal and quantitative sections of the GRE address even fewer of the objectives defined by the UTK coordinating committee than the ACT COMP exam or the McBer instruments.

The ACT COMP exam has two forms. the composite examination and the objective test. The composite examination involves the student in a range of activities that ask for not only brief written responses, but also short essays and oral communication. The objective test involves only multiple-choice questions. The former takes approximately six hours to administer and is expensive to score, the latter takes about two and one-half hours to administer and can be machine-scored. The relatively high positive correlation between the longer and shorter forms makes the objective test adequate for purposes of program assessment. Diagnosis and advisement of individual students on the basis of their scores are not primary objectives of the UTK testing program. The composite examination has been used by individual faculties, however, principally to assess writing skills.

The ACT COMP exam defines six areas that are at once measurable, basic to, and agreed upon in most definitions of attainment in general education (Forrest 1982). These are. (1) communicating, (2) solving problems, (3) clarifying values, (4) functioning within social institutions, (5) using science and technology, and (6) using the arts. The first three are broad skills that can be applied in the last three content areas. Questions are constructed in such a way as to measure a person's functional skills as well as his or her level of understanding in each content area.

These six areas hardly exhaust the issues and skills that have been included in a vital definition of general education. Not all of the basic skills (such as computation and foreign language), areas of knowledge, or attitudes and perspectives defined by the UTK coordinating committee in its statement of ideals are covered in the ACT COMP exam. Also not included are important goals such as lifelong learning and tolerance of divergent points of view. Some of these goals are not readily measurable, and some

are long-range competencies or broad perspectives that only manifest themselves over time. Thus the content of the ACT COMP exam represents only an agreed-upon core of vital areas in general education.

Experience at UTK has shown that, although faculty members are typically skeptical about the ACT COMP exam, a careful review of its design and goals, with recognition of its limitations, stimulates thinking about course and program design in ways that address a wider range of concerns than those usually defined by a single discipline. Workshops with faculty in which the ACT COMP exam is carefully assessed, including both the design and the process that shaped it, provide a new context for thinking about courses and instruction. Dimensions of problem solving, for example, or the articulation of values, can be made more explicit in any course if faculty members are committed to the endeavor.

Taking the ACT COMP exam is not the sort of experience that students usually associate with testing because it makes use of application and evaluation skills rather than recall and recognition. It is not a test for which students can formally prepare. It assesses the full range of the student's experience, skills, and knowledge gained in coursework (and elsewhere), and it tests the student's ability to deal with situations of the sort confronted by mature and concerned individuals. The use of news clips, art prints, scenes from films and dramatic performances, music, short essays, stories from popular periodicals, and sample letters to political and social agencies engages the student in a type of testing not often encountered in the classroom. The multiple-choice questions are skillfully designed to require the respondent to bring together all of his or her resources in creative and integrative ways. The activities that make up the test are selected with an eye toward encounters likely to occur in mature life and not toward the ability to do work at a higher level in a specific discipine.

### Administering the ACT COMP Exam

Since 1980, a sample of UTK seniors from each of the nine colleges granting baccalaureate degrees has been selected to take the ACT COMP exam just prior to the end of the spring quarter. In 1983, the testing of a sample of entering freshmen prior to the opening of the fall quarter was initiated. The freshman measure, coupled with the testing of a representative sample of seniors from all colleges, promised to yield valuable information on the achievements of students in general education and the value added by their work at UTK. This information could then be incorporated into broad program reviews, curriculum evaluation, and planning and budgeting processes.

In the winter of 1983, the UTK Faculty Senate approved a graduation requirement that seniors take part in at last one evaluation activity to aid the institution in the assessment and design of its programs and support services. Taking the ACT COMP exam fulfilled the requirement for a

selected sample of students, but the cost of administering and scoring the ACT COMP exam initially made it seem prohibitive to give it to all seniors. Thus some were asked to spend two hours taking an exam in the field, and all those not tested were required to take 20 minutes to complete a student satisfaction survey (see chapter 7). Many students were asked to take part in more than one activity, because in colleges such as nursing, engineering, and education almost all students took field exams as part of their professional program, yet a sample was needed to represent these colleges in the ACT COMP testing program.

In an institution as large as UTK, simply notifying those selected as part of the sample takes significant time and effort. Notification is carried out through the office of the dean of each college rather than through a central office because students generally feel strongest affiliation with their own college. Because the ACT COMP exam requires complex media arrangements and monitoring and cannot be given on demand, an appeals process has been established in cooperation with the college deans to handle schedule conflicts of students selected to take it. Generally, reassignment to another scheduled day or time solves the problem. The appeals process makes it possible to urge students who are taking field exams as part of their programs to agree to take the ACT COMP exam. It also ensures a representative sample from each college.

A student taking the ACT COMP exam finds it to be a unique experience. The exam demands that the student integrate a number of different types of skills and knowledge. Some UTK students said that it was "too easy"; others considered it "silly." Like some faculty members, students found it to be outside their usual experience. Indeed, an examination that is used for program assessment and not for individual evaluation is novel. As the ACT COMP exam is accepted among faculty members, it is desirable that they discuss it when advising students and in the classroom. If a balance between breadth of experience and depth of concentration is to be maintained, faculty members should pay more attention to advising students about the rationale for general education as part of their curriculum. Use of the ACT COMP exam may help overcome the perception of many that general education is something to be "gotten over" as soon as possible in order to get on to the "real business" of a college education—the major.

Experience in administering the ACT COMP exam and experience in motivating students, combined with the fact that the requirement is no longer new, have alleviated many of the problems initially encountered. The engaging nature of the exam and the staged way in which it is taken reduce attempts to falsify responses. The importance of the activity in determining the level of state funding for the institution and in assessing the design of curriculum and instruction is sufficient motivation for most students to answer the questions responsibly. Because the value of the results of the ACT COMP exam has been recognized at all levels within UTK, plans are being made to require all seniors to take it in order to graduate. In time, all freshmen may be tested as well.

## Results of Measuring Achievement via the ACT COMP Exam

It has been said that there are really three curricula in a college or university: (1) the curriculum described in the catalogue, (2) the curriculum that faculty members think they teach, and (3) the curriculum defined by what students learn. The ACT COMP testing program provides a means to link the three curricula and draw them closer together. The six areas covered by the ACT COMP exam have been developed based on a consensus about what is desired in general education in a variety of institutions. The ACT COMP exam measures actual student performance with respect to specific activities; that is, it tests students' knowledge based on their experiences in the six broad areas. Because the ACT COMP exam provides a bridge between broad ideals and actual student achievement, it suggests a context for designing and organizing courses in such a way that they transcend disciplines. ACT COMP exam scores provide evidence of the impact of reforms on curriculum, course design, and instructional techniques.

How will the information gained through the ACT COMP exam inform instructional development and assessment? The ACT COMP exam will hardly either change a century of emphasis on student development in a discipline or transform the discipline-based structure of an institution such as UTK. Disciplines and their manifestations in majors, departments, and divisions are not going to disappear. But the academy is more than a collection of discrete disciplines. It is a community with a vision of what constitutes quality education. The curriculum is the social contract that the community makes with the larger society; it is a statement that says what an education can and should do. It is imperative that the contract be articulated with specificity and precision and that the design of curricula, courses, and modes of instruction reflect its specificity and precision. The three curricula, in other words, should be drawn closer together.

Interest among faculty at UTK has risen because the ACT COMP exam is perceived not only as a source of data about how students perform, but also as a focus of thought about a range of goals that transcend restricted disciplinary interests. Such goals stand at the heart of rhetoric about the value of undergraduate education and the qualities of an educated person. To avoid the criticism of attempts at reform and renewal of general education, namely, "When all is said and done, more is said than done," it is vital that general-education values are cultivated in courses, programs, and instruction.

In five years of using the ACT COMP exam to test seniors, a clear pattern of scores has emerged across the six areas measured that is generally characteristic of all colleges at UTK. While it is possible to estimate freshman scores on the ACT COMP exam from their ACT composite scores and to compute from this the value added in relation to the senior scores, initial results of freshman testing indicate that freshman scores at UTK are actually somewhat below the estimations provided by ACT. Thus, when the

same individuals who took the ACT COMP exam as freshmen are tested again as seniors, actual value added will be greater than that currently being estimated.

The freshman pattern of scores across the six areas is also remarkably similar to that of seniors. Low entering scores in a particular area are paralleled by low senior scores and do not necessarily indicate undue weakness in a UTK program. The demographic information as well as the data on sequences of courses provided by freshmen and seniors when they take the test make it possible to determine which factors significantly affect overall scores, subscores, and score gain.

To date it appears that, while the curricula offered by the various colleges of UTK differ significantly, the overall ACT COMP exam scores and pattern of subscores are remarkably similar across colleges, and they are most directly related to the entering ACT scores of students in each college ($r = .50$ between entering ACT composite and ACT COMP total score at UTK). The greatest score gain is demonstrated by those with low entering ACT COMP exam scores ($r = -.41$ between ACT composite and score gain at UTK). As criteria for admission to UTK become more selective and entering ACT COMP exam scores rise, questions will occur about the amount and type of attention that will need to be devoted to general education.

The results of testing should be brought into the program review and planning processes of an institution. When new curricula are implemented, ACT COMP data will ideally play a major role in the evaluation process. Information about ACT COMP exam scores should be one of several factors considered by departments in the self-studies conducted for academic program reviews. Faculty willingness to make use of ACT COMP exam results and their confidence in them will be greater if they have previously found the issues covered by the exam to be vital general educational concerns that can stimulate course design and instruction. Using the ACT COMP exam as a resource for faculty development is thus always an option and need not await the accumulation of scores over a number of years.

Design and assessment of curricula take time, and reshaping of individual courses and patterns of instruction involve the efforts of many faculty members. In an institution like UTK, with a mission that calls for not only quality undergraduate instruction but also excellent graduate programs, research, and public service, competition for time is always intense. An initiative in performance funding that deals with quality of instruction must reward not only attainment but also improvement, which may involve experimentation and risk taking. The results of such efforts to improve program quality, especially in general education, are not likely to become clear for some time. It is difficult to assess and measure the impact of changes in curricula, course design, and instruction, especially over a four-year program of study. The discrete initiatives by many different instructors and departments accumulate only gradually to effect change in a total program.

Variable V, the general improvement component of the THEC Instructional Evaluation Schedule, acknowledges activities undertaken by an institution that improve the general-education component of its programs. Because significant changes are discernible only across time, annual benchmarks have limited meaning. When an initiative involves risk and fails, it merits acknowledgment nonetheless. The relatively small weight assigned by the THEC to Variable V provides minimal incentive in this regard.

## Conclusion and Recommendations

General education, particularly as operationally defined in the ACT COMP exam, is viewed by UTK faculty members and administrators as an essential component of undergraduate education. Most can agree that students should be taught to communicate, solve problems, and clarify values as ACT defines these skills. Thus the campuswide review of the THEC instructional evaluation criteria has produced more agreement with the concept of testing in general education than with any other standard of performance funding.

Selected faculty members in every college enrolling undergraduates have been given the opportunity to familiarize themselves with the ACT COMP exam. Those who have done so are intrigued by the design of the instrument and the nature of the questions asked. Two faculty development workshops have been devoted to further study of the exam and the rationale for its development. The outcome of these workshops is a group of faculty sensitized to the importance of higher order intellectual skills in teaching and the need to emphasize these skills when testing students.

There have been significant score gains on the ACT COMP exam from freshman to senior years. The pattern of value added has excited the interest of deans and faculty members who have seen the results. Further study of the correlates of achievement in general education may suggest experiences in course work or campus activities that will enrich the general-education experience for students (and increase ACT COMP exam scores). Although the ACT COMP exam does not measure achievement of all objectives set forth in the document defining general education at UTK, it is a useful starting point to which other measures may be added as time and resources permit.

The increased attention to the structure of general education described in this chapter is continuing at UTK, and it will intensify as the current plan to convert the university from a quarter system to a semester system by 1988-1989 is implemented. Deliberations about the change from curricula based on quarters to ones based on semesters will be informed by the results of the systematic measurement of competence in general education via the ACT COMP exam for freshmen and seniors.

The experiences of UTK with general education are not unique; indeed, they are probably representative of those of many institutions. UTK has been a pioneer, however, in instituting a testing program to

describe and evaluate its experiences. This is not to say that the ACT COMP or any other exam should be the only indicator of program effectiveness. Measures of affective growth and attitudes toward the educational experience must be included. The assessment process itself is beneficial because it focuses the attention of faculty, administrators, and students on general education, defines fundamental objectives for curricula, and provides important information for program review and planning. When the effort also results in a greater share of available funding, this clearly suggests that the priorities of the university and those of society are in line. The call to accountability is thus a meaningful challenge.

## References

Carnegie Foundation for the Advancement of Teaching. *Mission of the College Curriculum: A Contemporary Review with Suggestions.* San Francisco: Jossey-Bass, 1977.

Forrest, Aubrey. *Increasing Student Competence and Persistence. The Best Case for General Education.* Iowa City, Iowa. ACT National Center for the Advancement of Educational Practices, 1982.

Gaff, Jerry G. *General Education Today.* San Francisco. Jossey-Bass, 1983.

Tennessee Higher Education Commission [THEC]. "Instructional Evaluation Variables." Nashville, Tennessee, 21 November 1983.

Winter, David; McClelland, David, and Stewart, Abigail. *A New Case for the Liberal Arts.* San Francisco: Jossey-Bass, 1981.

CHAPTER SIX

# Measuring Achievement in the Major Field

*By William H. Calhoun*

THE GUIDELINES FOR performance funding developed by the Tennessee Higher Education Commission (THEC) for use by the state's colleges and universities mandate assessment of achievement in the major field. This may be carried out either through an externally validated test instrument approved by the THEC staff or through a locally developed test. Credit is awarded the institution if the program demonstrates improvement in either of these measures over successive testings, or if the program can show that its graduates exceed the performance of graduates of similar programs at comparable institutions.

This chapter addresses the manner in which the University of Tennessee, Knoxville (UTK), responded to the challenge of performance funding in terms of demonstrating levels of achievement by students in their major field, and illustrates the various issues involved in measuring this achievement. Testing competence in the discipline is an important criterion for demonstrating program quality and inevitably will be a factor in a system for awarding state funds on the basis of performance (Bogue and Brown 1982).

## Problems Associated with Measuring Achievement in the Major Field

Testing achievement in the major field is more difficult and more controversial than testing in general education. In essence, what one aspires to do is to establish a set of criteria for evaluation of a program rather than a norm based on relative rankings of students with respect to other students' performances. Such an aspiration should not be viewed as impractical, although it is indeed difficult to achieve. Attaining or improving quality in education, no matter how the process is defined, is an admirable goal toward which to strive (Pirsig 1974).

For general education, at least one national test is available that can be administered centrally with little involvement of individual faculty. For testing in the major field, few national exams are available. Faculty must be directly involved in assessing the appropriateness of a standardized exam for evaluating the achievement of local program goals, or in designing their own exam. If an existing test seems appropriate, the program faculty members should use that test at least once to determine how students in the major field perform. If there are national norms against which performance can be compared, and if the comparison is favorable, the minimum criterion for performance funding in the area of measuring achievement in the major field has been met. If a national test is not available or does not fit the program in question, development of a local test is necessary, and it can be conducted as suggested in the following section.

The THEC guidelines on testing in the major field are fairly straightforward, but there are some serious problems nonetheless. First, a program cannot show improvement over successive testings, as the THEC suggests, if it is presently doing an excellent job of educating its students. Second, if the testing is started and the program is adjusted to improve results, improvement cannot be shown for every successive testing. The students will reach the top of the scale and be unable to improve further (the ceiling effect). Even if the test has no theoretical upper limit, gain is reduced with each successive improvement in the educational program and will reach a ceiling at some point. Thus, continued successive improvements cannot be expected. In sum, the concept of showing improvement with successive testings may seem easy at first but, due to the ceiling effect, this approach cannot be used indefinitely. Yet, if external norms are not available or relevant, improvement data must be used.

A third problem has to do with the use of a standardized test. External norms may be difficult to obtain or irrelevant for local use. Where there has been some type of national program to assess knowledge gained in the major field, there may be useful norms. But for many national tests, the norms may not be useful because the standardizing group is not comparable to the group of students being tested in the local program.

Some professional fields provide examinations for purposes of licensure or certification. Since the content of these exams represents some consensus concerning concepts that "every professional in the field should understand," their validity for evaluation of local programs is increased. But the information about performance on these exams that is made available to the campus may not be useful for program evaluation. A dean may receive a mean total score for graduates who passed the test but no information about subscores. Scores in the specialty areas that the exam comprises are critical evidence of strengths and weaknesses in the curriculum.

Recent communications with individuals charged with administering state or national testing programs in architecture, engineering, law,

veterinary medicine, planning, and social work have revealed that some of the administering boards do not release exam scores (beyond percent passing) for institutions because they do not wish to foster competition among the institutions offering programs in the field. Some boards do not even collect sufficient descriptive information about persons taking the test to permit aggregation of data by institution. Some of the boards have responded positively to UTK requests for subscores, offering to make the information available immediately or to consider at a board meeting the request for changing the policy or adapting the methodology of collecting information to permit transmittal of institutional subscores. Other boards have indicated their intention to abide by existing policies that prevent release of institutional scores.

Regardless of the validity of a national exam for assessment of a local program or the availability of subscores, the norms provided for comparison of scores may not be appropriate. For maximum usefulness, norms for a given program should be based on scores obtained by graduates of peer institutions. The American College Testing Program (ACT) provides comparisons of mean score gains for groups of peer institutions on its test in general education. No other national testing program currently provides this service.

Finally, performance funding could develop into a fearful game of direct competition. Nowaczyk and Frey (1982) found that the performance on an assumed achievement test (the Graduate Record Exam [GRE] Advanced Test in Psychology) was more affected ($r = .65$) by the SAT scores of the students than by any other measures, including measures such as grade-point average (GPA) ($r = .44$-$.49$) or hours of psychology courses ($r = .30$). This is rather discouraging. Apparently, the best way to attain high GRE scores in psychology, thus indicating sound achievement in the major field, is to attract students with high SAT (or ACT) scores. If every college and university is involved in developing measures of achievement in the major, the competition for students with high scores on entrance exams could be fierce, and the extremes to which program administrators might go to show greater improvement than that of peer institutions could negate any benefit to be gained by performance funding.

## Conceptual and Operational Issues

Before the decision is made to use a test, conceptual and operational issues in measuring achievement in the major field need to be considered. Preparation in the major may play a rather minor role for some students, as in an individualized program that cuts across several disciplines, or it may play a very dominant role for other students, as in professional programs such as nursing, engineering, human services, and medical technology. For many professional and preprofessional programs, standard tests to measure achievement by seniors or graduates have been developed that can be used to demonstrate achievement of students in those fields.

The individualized major, in which requirements are minimal, and the professional program, in which requirements are quite specific, represent two ends of a continuum. For the typical undergraduate major, the number and specificity of requirements fall somewhere in between these extremes. For most major fields there will be common material that is studied by all students, and the faculty can settle on shared material as the domain in which students will be tested to evaluate achievement in the major field.

Many majors have such a common core, but there are options for specialization beyond the core. Thus, it may be necessary to develop different tests, one to measure achievement in the common core and one to measure achievement in the area of specialization. Examples of fields in which specialization is important are history and English. These are fields in which majors take a set of common courses and specialize in American or European materials. For these fields, if the multiple-test strategy is attempted, the task of measuring student achievement is much more demanding in terms of time and expertise, and it will be much more difficult than the expedient of using a single test that samples the field broadly. To achieve the task of obtaining a valid measure of achievement in the major field, the test must be tailored to the specific situation.

Involvement in a project that assesses program quality via student achievement in the major field may engender strong ambivalence in faculty members. This ambivalence would stem from the belief that a single definition of achievement (or quality) is narrow and specific and that the measurement techniques selected may overlook or leave out important aspects of the outcome of the educational process, such as development of values and the cultivation of student awareness about social issues.

The primary task of measuring achievement in the major is to focus on the specific kind of knowledge expected of students. For purposes of this analysis, the range of knowledge will be divided into four areas: (1) cognitive, (2) experiential, (3) affective, and (4) conceptual. Cognitive knowledge refers to the learning of general concepts in the major field. Largely factual in nature, this material can be assessed by a standardized, objective test in a straightforward manner. In the experiential area, the focus is on knowledge or skills gained through supervised practicum or laboratory experience. Examples include knowledge that is gained by students in human-services programs who are assigned a structured placement in a human-services setting, nursing students who engage in clinical placements in hospitals, and medical technology majors who learn a number of techniques for assaying biological systems. In this grouping, as well, are the skills acquired by studio art and music majors.

Measuring experiential achievement is not easy and usually requires subjective evaluation by supervisors or others of an art show, a musical recital, or a poetry reading. Subjective evaluation has some serious weaknesses, however. In the medical field, courts have disregarded the

clinical judgment of physicians and favored mechanical tests (Pinckney 1983). For some major fields, achievement of excellent performance is the most significant part of overall achievement in the major, yet developing procedures for measuring such excellence requires special attention and has pitfalls.

In the affective area, the reference is to attitude, emotional development, or change in students as they progress through a program. Nursing students, for example, are expected to develop certain attitudes about people and their health needs, as well as about the ethics of their profession. Students who take group-process courses aimed at helping them learn human-relations skills, such as special-education students and social work students, should be able to demonstrate achievement in the affective area. These achievements can be assessed by having students write essays or diaries about topics of concern that demonstrate their attitudes and how these change with experience. This is the most difficult area in which to show achievement via a multiple-choice examination.

Ethical views are normally assessed subjectively via peer review and interview or oral examination. Understandably, there are serious problems with the use of interview or oral examination data as the basis for conclusions about achievement. Nonetheless, for some majors ethical concerns are paramount, and they must be included in any valid measure of achievement in the major field.

For the conceptual or cognitive area, the measurement task is more readily accomplished than in experiential and affective areas because tests can be developed containing items that actually sample students' achievement. Questions need not tap only factual material, but can also sample the student's grasp of the conceptual aspects of the major field. Skilled test writers can test for conceptual knowledge.

The design of a comprehensive examination could easily reduce all areas to simple demonstrations of factual knowledge. This approach, however, clearly penalizes many students because major fields vary from being highly to loosely structured, and the task of measurement changes with the degree of structure. For some fields, such as English, music, and art, performance skills acquired through experience should constitute a major focus in measuring achievement. In other fields, knowledge is largely factual. For each case, a different type of measurement approach may be required.

### Existing Tests

An advantage of using existing tests is that they are usually standardized and accompanied by historical and normative data. Moreover, far less faculty time is consumed in selecting a test than in developing one.

For measuring achievement in the major field, the most widely known instrument is the GRE. The GRE is administered nationally by the Educational Testing Service (ETS), and has advanced tests for many fields

of study, including biology, computer science, literature, mathematics, physics, psychology, and sociology, among others. The GRE has long been in use, and until recently two sets of norms were available. From 1973 to 1977, ETS conducted the Undergradute Assessment Program (UAP) (Educational Testing Service 1978). Colleges and universities were offered the chance to administer the GRE to most or all of their advanced undergraduates (not just those bound for graduate school) and to use the results for purposes of program evaluation. Appropriate norms were thus provided for others interested in evaluating their programs. However, the UAP was not profitable and ETS discontinued it. The aging norms were declared invalid after June 1984. Existing norms for the GRE are based on annual test results from students who intend to pursue graduate studies and thus take the test voluntarily. This last set of data is obviously obtained from a selected group of highly motivated students, and results cannot be readily generalized to all college seniors in a major field.

Other tests that can be used include the College Level Examination Program (CLEP) tests. The CLEP tests are not widely used, however, because they were not designed for evaluating achievement in the major and do not provide the necessary norms for this purpose. A full listing of tests is available in Buros (1974). Several other national tests are commonly used, including the Engineer-in-Training Examination, the National Teacher Examination, and tests in other professional fields such as architecture, social work, planning, and nursing. Most professional programs have some type of qualifying test currently in use. The Medical College Admissions Test, the Legal Scholastic Aptitude Test, and the Dental Aptitude Test are among others that can also be used.

If an externally developed test can be identified, the next step is to decide whether the test fits the program in question. This determination requires some preliminary testing and correlation analysis. Many faculty members are not familiar with the techniques used to develop and evaluate tests. When this is the case, it is desirable to contact an experienced consultant to assist in the process. A content analysis should be conducted to show which areas of the field the test covers, does not cover, or covers incompletely. Adjustments in the test may be needed in order to bring the content into line with what students are expected to learn in the program. It should be kept in mind, however, that such tinkering voids existing norms.

The best approach is to use the test without modification on the first trial and then compare student performance to existing norms. If this comparison indicates that the students' mean exceeds the norms, and if faculty are reasonably satisfied that curriculum content is covered, the task can be considered complete. If the student scores do not exceed the norms, an item analysis should be conducted to detect specific student weaknesses. If students have not been exposed to certain areas covered by the test, or if they have been exposed to areas not covered by the test, some tinkering should be undertaken with an eye toward developing new norms and

readministering the test a year later. In this case, one might expect students to show improvement on the next administration of the test, thereby meeting one of the criteria of performance funding for evaluating achievement in the major.

The Department of Psychology at UTK elected to use an existing, standardized test—the GRE Advanced Test in Psychology—to assess achievement in the major. Measurement consequently focused on cognitive and conceptual areas of attainment. Disciplines that are biased in terms of experiential and affective components need to attend to those areas as well. Psychology is usually considered in a broad sense, and there is little specialization at the undergraduate level. Hence, a single test may be used to measure achievement. For fields where specialization is required or permitted, a second test with several options is needed to carry out a complete analysis of program effectiveness.

### Locally Developed Tests

There are both advantages and disadvantages to designing a local exam in the major field. Most important for purposes of program assessment and improvement, when faculty take the time to develop their own exam they have an intellectual stake in the results. They are more likely to use the findings to bring about change in curricula than they would be if standardized exams were used. The local test has greater face validity for the faculty who designed it, and it has greater content validity for evaluating the local curriculum. If a national exam provides evidence that students have done poorly, faculty members can rationalize that "the exam doesn't emphasize the things that we think are important." If the exam has been written by faculty members themselves, such a rationale does not justify poor performance.

Item quality and reliability (discussed below) may be lower in the locally developed test. But fortunately most campuses have faculty skilled in measurement techniques who can review local tests and suggest improvements. Such services may also be obtained from a consultant employed by a state or national testing agency.

The locally developed test also suffers from a lack of norms against which to compare students' scores. Thus even though the local test may be more useful for program evaluation and improvement, this advantage is offset somewhat because the scores obtained by program graduates can't be compared with scores obtained by graduates of comparable programs around the country. Yet even this disadvantage can be partially counteracted if two or more institutions agree to cooperate in designing, administering, and sharing the results of an exam.

To the credit of the developers of the THEC Instructional Evaluation Schedule, a program can receive as much credit for demonstrating improvement in scores over time as for establishing the superior performance of its

graduates on an instrument with national norms. The emphasis in the schedule on improving program quality is probably the principal justification for using locally developed tests. One can argue interminably about the technical flaws of tests constructed by faculty with no measurement expertise. However, as the test results are aggregated and used to make judgments about the quality of programs rather than the relative competence of individual students, the importance of the technical problems of the tests is minimized and the ultimate outcome—program improvement—is essentially positive.

### Constructing a New Test

Constructing a new test is a challenge. Most faculty have little experience in test construction and find out quickly that the task is not easy. Fortunately, there are numerous aids for undertaking test development. A useful source is *How to Measure Achievement,* by Morris and Fitz-Gibbon (1978). The booklet includes a set of excellent references, such as *Evaluation in Geography: A Resource Book for Teachers* by Senathirajah and Weiss (1971), and it systematically takes the reader through the entire process.

In some fields, there are sets of test items that have not been copyrighted, meaning that they can be used without permission. Most publishers provide these free of charge when they are requested directly and accompanied by an explanation of the intended use. In psychology, test files are available for the standard textbooks in the general course. Some of these files are written in a general manner so that they can be used with any textbook, for example, *Test Item File* to accompany *Psychology* by Scott, Foresman and Company (Costin, Slaw, Ory, and Landesman 1983). Publishers may have these questions available on computer disc, thus making local test construction much simpler than it would have been just a few years ago.

Constructing one's own test requires time and effort. The initial cost of the task is great in terms of faculty and secretarial time, but since the test may be used again and again, the cost decreases over time and may work out to be less than that required by ETS to administer the advanced GRE. If this type of cost assessment is included in the overall analysis, it may be determined that the cost of test construction is offset by the economies achieved with continued usage.

When a department decides to construct its own test, the first issue to be settled is the test format. The test may be composed of objective items, either true-false or multiple choice, or subjective items, which could include short-answer or essay items requiring longer answers. If the decision is made to use an essay examination, problems regarding the validity of the exam will arise. However, it is useful to note that the Bar Exam is an essay examination and that the legal profession has been able to defend its use in court. Essay exams require considerable effort to grade, and often

evaluation standards are not clear. Since this type of examination will be open to challenge in terms of content and method of scoring, one must have willing and skilled graders.

A second issue to resolve is the content of the examination. This can be arrived at in many ways, but in every case the faculty should be included in the decisionmaking process. Faculty members should be polled regarding their ideas about studies and textbooks from which items should be taken that define the field.

Issues in test construction that need to be considered are largely matters of reliability and validity. These issues have been addressed in detail in many references, including the Morris and Fitz-Gibbons reference, and will not be considered in detail here. However, brief reference to the reliability and validity issues is necessary. Reliability refers to the notion that any given test should yield consistent results. Normally a test will be reliable if it samples the material in a fair manner and if the material is relevant for the student being tested. Reliability can be estimated in a number of standard ways. The simplest technique is to administer a single test and apply a special formula that estimates reliability by splitting the test into halves and computing correlation coefficients between the halves. For the test to be considered reliable, these coefficients should be in the neighborhood of .80 to .90.

Validation of a test is less straightforward. The simplest notion of a valid test is that it measures what it purports to measure. For example, a new thermometer is valid if its measurements correlate well with those of a standard thermometer. For a paper-and-pencil test, the problem of validity is quite serious because there are few standards with which these tests can be compared. For example, the principal method for validating a new intelligence test is to relate its scores to an accepted intelligence test such as the Stanford-Binet or Wechsler Adult Intelligence Scale. However, even these tests do not possess perfect validity. Hence, using them as standards will not give true validity estimates. Another way to establish validity is to use some terminal behavior as the major criterion, such as GPA at graduation or performance on the job. These measures are not entirely valid in themselves, and do not provide a stable criterion against which to validate the examination. If we consider these issues in terms of measuring achievement in the major, there is no standard with which every test of achievement can be compared to estimate validity. Thus, there is no easy solution to the validity problem. Given the nature of the concept being measured, each test must be validated as best it can be.

For the newly constructed test, one needs to strive for as high a level of validity as possible. The easiest type to establish is face validity. A face-valid test is one in which the persons who review or take it agree that the items in the test are proper for the specific field being tested. In other words, the test makers and test takers believe that the items on the test are

items that students should know. Face validity can be enhanced if the major teachers in a program have input into the test content and students have a chance to affect the final test content in some way.

A second type of validity that is relatively easy to establish is content validity. This concept is an extension of face validity, and it concerns the appropriateness of the content of the examination to the field. If the items are considered relevant to the field by the faculty members and students, the examination is content valid.

The locally constructed test probably will be more face valid and content valid than will the GRE advanced test in the field. Moreover, if faculty and students are involved in its construction, their attitude toward the test may be much more positive than toward the GRE. Finally, the test can be administered under more humane conditions than is the case for the GRE.

## Conclusion and Recommendations

Measuring achievement in the major field can have substantial benefits. First, students are positively affected when someone shows interest in their achievements. Second, department faculty are sensitized to areas of strength and weakness in the program. Faculty members typically grumble, but most will make changes in their syllabi and in the methods they use to teach the material. Both of these factors can have a positive impact on students' learning in their major. The most important outcome is that faculty members responsible for instruction in the major field gain information that they need in order to improve the program for future generations of students, which may mean changing the types of courses offered, their content, and the rigor with which they are taught.

The THEC has not addressed the methodological issue of how to equate groups with respect to ability and other relevant variables when student scores obtained in the first five-year cycle of performance funding are compared with those obtained in the second cycle. Test-score gains (or losses) should reflect changes in program quality rather than changes in the characteristics of the two student groups whose scores are being compared. Moreover, faculties that have invested time in developing an exam will want to use it frequently, revising items and updating content continuously to improve the technical characteristics of the instrument. Will the form of the test used five years hence be equivalent to that used initially?

The problem of providing quantitative evidence of program improvement using students' test scores is one very important concern regarding the THEC standard of assessment in major fields. The other concern, which is even more important, is that a set of scores on a single comprehensive exam simply is not a justifiable means of evaluating some programs. For programs oriented toward performance, such as art, music, dance, and theatre, a juried performance constitutes a more valid comprehensive measure of achievement than does a paper-and-pencil test. Yet

each performance must be judged individually, and the averaging of ratings for purposes of reporting a single score to the THEC requires a leap forward in the technology of measurement that few are prepared or disposed to make. The same objection is being applied to the aggregation of scores on individual comprehensive exams given to candidates for master's and doctoral degrees.

A solution to the foregoing measurement problems is to give much greater weight than is currently the case in the THEC Instructional Evaluation Schedule to the comprehensive program review process as the quintessential instrument of academic program evaluation.

## References

Bogue, E. Grady, and Brown, Wayne. "Performance Incentives for State Colleges." *Harvard Business Review* 60 (November/December 1982). 123-28.

Buros, Oscar K., ed. *Tests in Print II.* Highland Park, N.J.: Gryphon Press, 1974.

Costin, Frank; Slaw, Kenneth M.; Ory, John C.; and Landesman, Ann. *Test File to Accompany Psychology.* Glenview, Ill.. Scott, Foresman and Company, 1983.

Educational Testing Service [ETS]. *Undergraduate Assessment Program Guide.* Princeton, N.J.: ETS, 1978.

Morris, Lynn, and Fitz-Gibbon, Carol Taylor. *How to Measure Achievement.* Beverly Hills, Calif.: Sage, 1978.

Nowaczyk, Ronald H., and Frey, Jerry D. "Factors Related to Performance on the GRE Avanced Psychology Test." *Teaching of Psychology* 9 (October 1982):163-65.

Pinckney, Edward R. "The Accuracy and Significance of Medical Testing." *Archives of Internal Medicine* 143 (March 1983):512-14.

Pirsig, Robert M. *Zen and the Art of Motorcycle Maintenance.* New York. Bantam Books, 1974.

Senathirajah, Nallamma, and Weiss, Joel. *Evaluation in Geography. A Resource Book for Teachers.* Toronto, Canada. Ontario Institute for Studies in Education, 1971.

# Measuring Perceived Program Quality

*By Kent D. Van Liere and William Lyons*

WHY STUDY OPINIONS about program quality? The most important indicators of program quality are structural features and outcomes of the program. These provide the most obvious indicators of program success or failure. They do not, however, reflect the shifting and illusive thinking that characterizes opinions and perceptions about academic programs, nor do they indicate the underlying reasons for the success or failure of a program. That students fall below the norm on a licensing exam begs the question, "Why?" The single most important reason for studying perceived program quality is to obtain information for designing program improvements.

The evaluation of program quality can take many forms. Preceding chapters have focused either on evaluation of structural features of a program, such as accreditation, or on the outcomes of a program, such as test scores of graduates. In this chapter, methods for evaluating the quality of academic programs by using survey research are examined. Specifically, this chapter examines opinions about program quality as expressed by students, alumni, or other relevant groups, as well as the methods by which these opinions can be assessed so that they provide useful information for improving program quality.

The purpose of this chapter is to outline several issues that have been identified in the process of measuring student perceptions of program quality at the University of Tennessee, Knoxville (UTK). This discussion provides an assessment of the utility of measuring perceptions for performance evaluation and funding. The aim is to construct an information base that can be useful to administrators and faculty members who are in charge of classroom experiences and programs leading to a major. This focus differs substantially from existing approaches that tap student perceptions of university services, such as library and health-care services. Thus, it is as important to explain the general conceptual approach as it is to discuss methods and results.

### The Research Design

The variable that addresses perceptions of program quality in Tennessee's Instructional Evaluation Schedule includes two standards. First, one-half of the total credit can be achieved by conducting a representative evaluative survey of enrolled students, former students, or community members/employers. Second, equivalent credit can be earned either by surveying two of the above groups simultaneously or by surveying the same group across time. In both cases, maximum credit is earned only if the institution demonstrates that it has used the results to formulate and implement improvements.

The two assumptions that provided the general parameters within which the research has been carried out have been interpreted as follows in the research design:

- The dominant methodology should be survey research
- The survey should produce results amenable to recommendation of specific actions
- The survey should focus on academic programs and instructional improvements

Given these guidelines, the assessment of student opinions provides an enrichment of objective measures of program quality such as accreditation and scores on exams in the major or in general education.

A central question is: What should be measured? The guidelines of the Tennessee Higher Education Commission (THEC) for performance funding provide a range of possibilities. The discussion of students' opinions often focuses on their satisfaction with programs, suggesting that the degree of satisfaction is an appropriate area to query. The methods available to measure satisfaction with programs are, however, generally unsatisfactory.

The concept "satisfaction" is used widely in reference to evaluation of a program, but it can be confused with "perceived quality." Satisfaction implies more than a mere evaluation, and in many ways it is quite different from the notion of quality. Satisfaction refers to both students' expectations about a program and their evaluations of it. When students reply that they are very satisfied, the response can mean either that they expected high quality and received it or that they expected low quality and received it.

Is it more important to produce satisfied students or students who perceive their programs to be of high quality? Each objective is important in a different way. For purposes of short-term policy adjustments, which from the university's point of view are the *raison d'etre* of the survey process, evaluations of the quality of services are the most useful. Entering students with very low expectations about a program's quality may be quite satisfied if the program turns out to be better than expected, nonetheless, they may still feel that the program is not very good. Thus, the appropriate focus is perceived program quality rather than student satisfaction.

The second issue concerns the types of programs that should be evaluated. This focus was partially influenced by a review of existing measures of student satisfaction that was carried out by a faculty task force examining the use of information on student outcomes. As part of the project sponsored by the National Center for Higher Education Management Systems (NCHEMS) and the Kellogg Foundation, this task force was charged with the responsibility of identifying appropriate methods to assess student satisfaction with programs and services. Several existing instruments were examined, including the ACT Student Opinion Survey (SOS). The committee concluded that all of the instruments suffer from the same limitation. they provide very little of the kind of detailed information that is needed by those in charge of academic programs. In general, the surveys focus on student satisfaction only to the extent that they include questions about services provided to all undergraduates.

The specific locus of satisfaction is difficult to isolate, but there is reason to believe that students' feelings about academic programs are influenced by at least two additional factors. These are experiences with the major program and experiences in the classroom. Thus, attention was focused on three levels of student experience with the university. (1) general services, (2) the major program, and (3) classroom experience.

The third issue involves the sampling plan for the surveys. This focus is influenced by two research objectives. The first is to provide information that is representative of the whole university, and the second is to gather information that can be used by department heads and other administrators in charge of academic programs and classroom situations. To achieve both objectives simultaneously would have required a very large sample size and considerable cost. This issue was solved by using a combination of samples that provided aggregate information for the university as a whole, as well as detailed information on a limited set of departments (described in detail below). Thus, the general information needed for monitoring universitywide programs was provided, and a few departments were supplied with the kind of specific information necessary to make program changes that would ultimately improve program quality.

### Instrument Design

This section describes the process developed to survey currently enrolled students regarding their perceptions of program quality. Previous efforts to measure student opinions of programs at UTK used existing measures of student satisfaction that were commercially available. These efforts were successful in meeting the requirements of performance funding as specified by the THEC, but they were not very useful to the university. For example, the survey conducted in 1982 dealt with only general university programs and services. Information was collected about the library, registration, the health service, and other services available to the campus population. While potentially valuable, this information did not provide

the kind of feedback that could be used to improve instructional programs. To overcome this problem, a decision was made to design an instrument that provides more specific information.

The design of the instrument involved a series of three groups of items (see appendix D, tables 2, 6, 9, 12, 15, and 18). First, student reaction to general services was measured. A variety of services was included in the questionnaire. Students rated each service according to four questions. (1) How frequently do you use the service? (2) Why don't you use the service more? (3) How do you rate the quality of the service? and (4) How important is it to you that the university seek or maintain excellence in this service area? These questions provided information on universitywide services that was quite similar to that contained in other broad surveys of student satisfaction.

The second group of questions focused on experiences with the major program. Students were asked to identify their intended or present major and rate a range of services related to that program. Program services or features such as advising, availability of courses, opportunities for interaction with faculty, and the quality of special events were included. Students rated each item in terms of its quality (excellent to poor), and they were asked about its importance (How important is it that the program seek excellence in this area?).

The difficulties encountered in constructing the second set of questions were of two types. First, how should students with no major or more than one major be handled? Students who had no major or no intended major were simply asked to skip the section. The students with more than one major were free to choose the program they wished to evaluate. The second problem involved the choice of items. Major programs differ in substantial ways from each other. Some involve lab work, for example; others require internships or emphasize writing. No single set of questions effectively taps all these dimensions for every program. Constructing different items for every program was too costly, however, and eliminated comparability. Items were therefore developed that generally fit most programs and provided a "not applicable" response for the exceptions.

The third group of questions dealt with classroom experiences. This set of questions is similar to items included in the evaluation form that an individual instructor asks students to use at the end of a course. It measures perceived quality and importance of achieving excellence in areas such as course content, instructor's teaching style, course objectives, and availability of the instructor. As noted in more detail below, students were asked to rate a specific class chosen randomly from their current course schedule rather than rate specific items for courses in general. Additionally, they were asked whether the class was required for the major, recommended for the major, or an elective. They were also asked whether the course was taught primarily by a faculty member or a graduate student.

A final section was included that tapped students' overall satisfaction with the university, their social life at the university, and their academic experiences. Although "student satisfaction" had not been selected as the main concept to be measured, it was useful to include questions that allowed the researchers to relate any component or subcomponent of the questionnaire to social, academic, and overall satisfaction with the college experience.

### Sampling Design

The major objective of the research design, as noted earlier, was to construct a study that would produce survey results that would be generalizable to the whole university and, at the same time, useful for department heads and deans. If it is assumed that the most important focus for instructional improvements is classroom experience in the major field, then it becomes critical that the information is collected in such a way that it can be broken down by college, department, and program. A sample of 500 students would provide estimates of perceived quality of the university as a whole, but it would not permit analysis of perceived quality of the college. Expanding the sample size to 1,000 would provide enough students for college-level analysis, but this would not allow breakdowns by department. An even larger and more costly sample would be required to achieve breakdowns by department.

The surveys completed during the spring of 1983 used three separate samples to address the above concerns. First, to ensure that the data would satisfy the THEC guidelines and provide a baseline for comparison, a representative sample of all full- and part-time undergraduate students was taken. This sample was stratified by college, and small colleges were oversampled to ensure that at least 100 students from every college were included. This sample included 1,155 students, and their responses were used to assess the level of perceived program quality across the university. These data were also used to measure perceived program quality for each college, as well as to compare colleges to each other and to the university as a whole.

The second and third samples were designed to provide data useful to department heads. In addition to the universitywide sample, five departments were chosen for more detailed study. Since the effort was experimental, the departments were not chosen at random. Rather, the cooperation of a range of types of departments was solicited. For each department, two samples were drawn, including a sample of majors and a sample of students taking classes in that department during that quarter. As a general rule, 300 students per department were sampled. The usual break down was 100 majors and 200 students from the classes. These numbers were adjusted to accommodate differences in the various departments.

An important dimension of the design for the sample of students in a department's classes was that the student was asked to evaluate a specific

class, not classes in general. The instrument provided an opportunity for an individual class to be identified for each student to consider in completing the classroom-valuation section. For the university sample, the class listed on the questionnaire was randomly chosen from the list of all classes that the student was taking that quarter. For the departmental samples, the class was randomly chosen from the list of classes that the student was taking in that department. These departmental samples were used to develop a report for each department head that summarized the evaluation of his or her program and compared it to that for all programs in the college and university. Additionally, the report showed the evaluation of classes offered by the department and compared this to the evaluation of all classes in the college and university (see appendix E).

The data were collected between April 25 and June 8, 1983, a period that encompassed the last five weeks of the spring quarter. The questionnaire was distributed through the mail to the student's current residence. An initial mailing and two follow-ups were used to elicit responses. The first mailing included the questionnaire, a cover letter from the chancellor, and a postage-paid return envelope. The second mailing was a postcard reminder. The third mailing included another copy of the questionnaire, a cover letter, and a return envelope. The procedure basically followed currently accepted practices for mail-survey research. Of the 1,155 students in the random sample of all university undergraduates, 809 returned usable questionnaires for a response rate of 70.0 percent. The response rates for each of the departmental samples were approximately the same.

## Outcomes

The survey process described above resulted in two primary products: a report summarizing the responses for each college and the university as a whole, and a report to each department head summarizing the responses of students in the department and comparing these responses to those for the relevant college and university. Examples of each report and some of the results will be discussed in this section.

The report based on the universitywide sample provided measures of perceived quality across programs and colleges, and was thus useful as a baseline for a variety of comparisons. Tables 6, 12, and 15 in appendix D are examples of the comparisons made among colleges at the university as a whole. Table 6 represents the ratings of perceived quality for general university services. For the administrators in charge of major university services, the data provided both a relative rating of their particular service as compared to others (as in the first column in table 6) and an indication of the differential ratings given by students in the various colleges (as in any row of table 6). Similar breakdowns were also created using other background characteristics of the students such as classification and grade-point average (GPA), as well as other factors like frequency of use of a particular service.

Tables 12 and 15 of appendix D are comparisons of ratings of perceived quality of various components of the major and classroom experience across colleges. These results are of most interest to department heads and deans. For example, table 12 indicates that the availability of a student's advisor (the first row) is rated much more favorably in some colleges than others (means range from 2.5 to 3.2). This type of result provides important feedback on the perceptions of students in these academic units. The researchers were also able to analyze these questions by student classification and GPA.

The second type of report was developed specifically for the departments in which further study was conducted. Tables 1, 2, and 3 in appendix E provide examples of the kinds of information these reports contained. Table 1 shows the ratings of quality of classroom experiences given by students taking classes in the department during the spring quarter. The table shows both the percentage breakdowns for each item and a comparison of the department's average rating with that of the relevant college and the average for the university. Table 2 reflects the ratings of various components of the program by students majoring in the department. These ratings are also arrayed to show both the percentage distributions and the comparison to the relevant college and university. Finally, using the background information included on the survey instrument, a profile for each department head of the types of students majoring in the department (table 3 is an example) was constructed. In the reports to the department head, no formal interpretation of the results was provided. Only the basic information about how the data were collected was made available.

### Where Should Policy Changes Be Focused?

An important series of observations could be made because the survey was administered to representative samples of the university community as a whole as well as to each college. If the colleges are sufficiently different from each other, is it appropriate to make a wider generalization to the entire university? Because five departments were singled out for separate analysis, this allowed us to speculate on the appropriateness of generalizations to the college level. The most meaningful variation to be found, it was expected, would be among departments within colleges.

Analysis of the responses to the survey indicates that, in fact, the meaningful locus of variation shifted from university to college to department, as general university services, programs associated with the major, and classroom performance were considered. The results suggested, for example, that there was no meaningful variation among the colleges as to how teaching was perceived (see appendix D, table 15). There was variation, however, among the five departments that were chosen, and this variation was substantial. With respect to the classroom experience items, one department ranked well above universitywide averages and those of its college, while another department ranked well below its college and the

university. These results allowed policymakers to observe the factors that separated departments in which teaching was positively received from those in which it was negatively perceived. Interpreting these differences is more difficult. Perhaps departments that scored well simply had individuals more skilled in teaching. This would put the meaningful variation at the individual level. Or there may be something about the subject matter, the size of classes, or the physical environment for learning that is fundamental to the variance among departments. The latter interpretation directs attention to departments as departments, rather than to departments as collections of individuals. In any case, the data suggest that there is nothing about colleges as colleges that would indicate a need to enlarge further the focus of research. In other words, to delve more deeply into teaching issues would necessitate the study of classes within departments.

Unlike the case involving teaching, the provision of departmental services connected with the major showed meaningful variation at both the department and college levels (see appendix D, table 12). That is, those factors associated with the department as a whole, such as advising and curriculum, varied both within and among colleges. Any policymaker whose interest is in maximizing positive reactions to this plane of university life would therefore be advised to focus attention on college policies as well as department policies. At UTK, the College of Agriculture, for example, was particularly successful at creating a more personalized climate in which students knew more professors and had access to their advisors. Other colleges were notably less successful in this regard. The variation suggests the possibility of viewing a university as a laboratory of sorts, with the successful experiment of one college serving as the model for another.

Finally, there was no substantive variation among colleges in students' evaluations of universitywide services such as the library or counseling services (see appendix D, table 6). Indeed, this was the only component of the evaluation instrument that it made sense to address in terms of general student reaction without differentiating by college or department. The task confronting policymakers in this area, while not necessarily easy, is not complicated by the existence of a different model for success in each college.

## Conclusion and Recommendations

Has the THEC standard incorporating referent-group surveys achieved the desired end of improving the quality of education at UTK? It is too early to make any conclusive statements, but the likelihood of improvement in the quality of the product offered by the university has certainly been increased. Decisionmakers in the various units involved in the survey now have access to information they did not have previously. At this writing, several departments or colleges are attempting to follow through on the survey results. These efforts range from further study to actual program changes.

Was the process cost-effective? Certainly, when the financial rewards associated with performance funding are considered, the university's answer is a resounding yes. The cost of the 1983 student survey was approximately $14,000, an amount that includes one-time costs related to instrument development that need not be repeated in subsequent years. Given the university's annual budget, the THEC award for a single survey under Instructional Evaluation Variable IV was $290,000 in 1983-1984. Thus, in simple dollar figures, the return on the investment at UTK was better than 20 to 1.

Is surveying a good method to measure quality? This is a difficult question to answer. Surveys are useful in diagnosing strengths and weaknesses, particularly when appropriate comparisons can be constructed (such as our comparisons of colleges and/or departments). These comparisons are useful primarily because the questions are asked of everyone in the same fashion. The meaning of the answers is more elusive. Survey researchers have warned repeatedly of the pitfalls in the wording of questions and response formats. Slight changes in wording can totally shift the meaning of an answer to a question on perceived quality.

It probably does not make sense to base performance funding on the absolute improvement in perceptions of quality of a program. What is important is whether survey results are leading to program changes that improve the quality of the program. If such changes are made, they will ultimately show up in the structural features of a program, such as accreditation, or in the outcomes of the program, such as student scores on licensing exams. These are better indicators of the quality of the program relative to its competitors in the educational milieu. But limited experience indicates that the surveys, when properly developed, conducted, and reported, can play a key role in instructional improvements. In regard to Variable IV of the THEC criteria, this indicates that the language of the variable should emphasize the use of survey results to attempt program improvement rather than actual shifts in the perceptions of quality.

This approach suggests that action with the intent of improving the quality of academic programs should occur at college and departmental levels, particularly the latter. Based upon initial findings, these conclusions are worth considering. First, it is difficult to justify a universitywide approach to some of the most critical problem areas. Second, the success of surveys like the one described in this chapter depends largely on the collection and dissemination of data that can be properly used at the college and even the departmental level. Finally, the whole enterprise is worth doing only if deans and department heads are willing to make use of the information. Given these caveats, it is safe to say that the survey methodology described in this chapter has yielded more information, aroused more interest, and led to more immediate adjustments in programs than has measurement in any other outcome area. It is therefore recommended that this component be increased in significance in the THEC Instructional Evaluation Schedule.

CHAPTER EIGHT

# Assessing the Quality of Higher Education Through Comprehensive Program Review

*By Mary P. Richards and C. W. Minkel*

ACADEMIC PROGRAM review is the systematic evaluation of a degree program for the purpose of improvement, continuance, consolidation, or elimination of that program. Although a few institutions have closely monitored academic programs for decades, many others, both public and private, were compelled by financial exigency and external influences to adopt formal review procedures in the early 1970s. Many state agencies and coordinating boards for higher education had been created by that time, and they had been given the power to mandate accountability for the quality of education from state-supported institutions. What began as an institutional tool to monitor quality and indicate directions for improvement ultimately attracted the attention of state governing boards as a potential means of assessing the quality and efficiency of their member institutions. Currently, approximately two-thirds of state systems of higher education or coordinating boards utilize program review, and one-third of all colleges and universities maintain review procedures (Barak 1982). Initiatives such as performance funding, started by the Tennessee Higher Education Commission (THEC), draw on program review as one means of evaluating institutional performance. The purpose of this chapter is to consider the effectiveness of utilizing program review to assess the quality of higher education.

## Approaches to Program Review

Numerous approaches to assessing the quality of higher education have been tried over the past two decades. However, assessment by program review has now been widely accepted because, when carefully designed and conscientiously implemented, it encompasses and surpasses other approaches. Reputational studies, such as the Roose-Anderson and Cartter reports (which rank graduate programs), have been found by

academic administrators to be useful but limited as a means of assessing quality. These reports attempt to assess the effectiveness of a given program and the quality of its faculty according to opinion of experts in the field. However, a number of objections to this procedure have been raised. First, many small institutions are omitted from these reputational surveys. Second, only a selected number of graduate programs are considered, and most of these are the traditional academic specialities. The much larger undergraduate offerings are omitted entirely. Third, reputational surveys tend to lag behind the current status of an academic program, it may take at least a decade to gain or lose a reputation for excellence.

Systems analysis as described by House (1982) has been tried in one form or another by several state governing boards in order to achieve increased efficiency through evaluation. This approach assumes that the goals of academic programs can be identified and that attainment of goals can be measured. As House observes, however, "The difficulty with the systems analysis approach is that educational programs seldom lend themselves to being measured by a few simple quantitative outcomes' (p.8). Although quantitative measures can be usefully included in a thorough program review, they are not sufficient to capture the complexity of a multigoal academic program.

Related to the systems approach to program evaluation is the value-added approach Students take standardized tests at the beginning and end of their academic careers, and the difference in performance indicates what the students have learned or the value added by experience at the institution. The information that this approach produces is limited to the skills or knowledge measured, and thus it provides information about only a small part of the overall development of a student. Moreover, if a teacher is motivated by the tests rather than the material, the value-added approach can lead to teaching only the facts that will be asked on the test.

Another approach that provides only limited information about program quality is the assessment of resources, including size of library collection, adequacy of physical facilities, and availability of up-to-date equipment. Assessments such as these are often included in accreditation reports and should also be included in a comprehensive program review. They are particularly useful in the assessment of graduate programs. Nonetheless, they are insufficient for a comprehensive evaluation of program quality and effectiveness.

The approaches described above lack the kind of subjective, overall assessment that is available only in peer review. Judgments about program quality, if they are to be useful, must be made in relation to resources available, complex goals, and performance outcomes. Outside experts—peers in the field—have much to offer when comparing the program with others around the country, but peers located within the institution are also needed to assess the program from the perspective of the institution itself. The human factor in peer review provides the subjective, holistic assessment of quality and effectiveness that cannot be obtained from other

approaches. An institution that hopes to improve its programs should consider peer review as the heart of the evaluation process. State agencies that plan to increase the return on investments in higher education should consider peer review as a primary source of information about program quality.

### Problems with Program Review

Since program review has so many positive attributes, one would expect coordinating agencies to use it as the principal means of assessing program quality at their member institutions. Yet in Tennessee, when the THEC was developing criteria for performance funding, program review was assigned only a minor role. Variable II in the THEC Instructional Evaluation Schedule accords institutional program review only one-third of the credit accorded objective testing in major fields of study.

Certain factors affecting program review made it less attractive to the THEC than other means of assessment. First, because the purposes of program review can differ depending upon the source of the mandate, some review processes may not directly address the issue of quality. The goals of an accreditation review, for example, may not mesh with the needs and objectives of a given institution. Some types of assessment that are mandated externally emphasize efficiency, acountability, reallocation of funds, and termination or reduction of programs to the exclusion of quality. As Wallhaus (1982) has shown, a state's objectives for program review typically include the following. (1) the development of a statewide master plan encompassing policies and priorities, (2) the elimination of unnecessary program duplication, and (3) the identification of needs for new programs. On the other hand, an institution's objectives for program review include the provision of assistance with (1) institutional planning, (2) allocation of internal resources, and (3) monitoring to ensure excellence. The chief problem for the THEC is that program-review processes are very difficult to compare with one another.

In addition to its focus on accountability, statewide program review has a second more serious limitation. when a uniform mechanism is applied to a variety of institutions, the unique mission of each may be overlooked. For the sake of comparison, institutions may be asked by state agencies to use objective data, which, as has been noted, give only a slice of the total view of the nature and quality of an individual program within a given institution. These types of measures indicate where an institution or program ranks in terms of grade-point average (GPA) of entering students, but they do not take into account the goals and priorities of the institution or program, nor do they offer suggestions for general improvement. To achieve both the statewide and institutional goals, the process of program review must be based in the institution and tailored according to the particular mission of the institution. At the same time, it must be capable of generating the kind of information needed for statewide planning and accountability.

## The Relationship of Program Review to Planning and Budgeting

Program review has much to offer the internal planning process of an institution. In fact, to be effective, program review must be related to the regular planning and budgeting function of the institution. Program review is often not implemented; reviewers' reports are received and filed, but these in themselves do not provide the ways and means for implementation. A review may therefore be perceived by faculty and administrators as simply an exercise without results, and consequently the process has relatively few benefits for the participants or the institution. To be demonstrably effective, program review must result in change that is brought about by administrators' and faculty members' responses to the recommendations contained in the final report. Most recommendations require funds for implementation, and they must therefore be considred in the planning and budgeting process of the institution.

The issue of determining the effectiveness of institutional program review is worth considering in greater detail. The self-study document, a necessary preliminary to any comprehensive review, sets forth the unit's goals and supporting evidence to indicate how well the goals have been achieved. This document can serve as the basis, or benchmark, against which subsequent achievements can be compared. Then, in the course of peer review, the reviewers can conduct their own investigation of program quality and effectiveness and make recommendations for improvement.

In the final phase of program review, the recommendations are considered and accepted, rejected with cause, or modified to support broader institutional objectives. The ways in which departments, colleges, and institutions respond to recommendations and improve their programs can then be demonstrated through evidence such as improvements in standardized test scores, new or reallocated funding to bring equipment up to date, improvements in the results of objective teaching evaluations, increased volume of external support for research, and growing numbers of articles published by faculty members in refereed journals. These are the types of follow-up actions that can be measured and documented by departments, colleges, and institutions to demonstrate effectiveness of program review.

## Use of Program Review in Performance Funding

When the THEC adopted its current Instructional Evaluation Schedule in 1981, the University of Tennessee, Knoxville (UTK), had just expanded its program reviews to focus not just on doctoral programs, but to encompass entire academic programs. The new approach includes the following. (1) a self-study component, (2) interviews of faculty, students, and administrators by a team of internal and external consultants, (3) formal reports with recommendations, and (4) scheduled follow-up meetings to address the recommendations and, where appropriate, incorporate them

into the institutional planning and budgeting process. The internal consultants represent the college, the Undergraduate Council, and the Graduate Council. They are selected from faculties and heads of related departments. External reviewers are leaders in the given field from other institutions in the region or nation.

In response to Variable II of the THEC schedule, which states that evidence of value added must be provided for students who progress through an undergradute program at the institution, the component "Indicators of Program Quality" has been added to the criteria for the self-study document that is prepared in advance of the review. Indicators may include assessments of the program by enrolled students and/or assessments obtained by standardized tests such as the Graduate Record Examination (GRE) advanced test in a particular field. Participants in programs under review are also asked to state desired outcomes for students so that goals and results can be compared. The portions of the self-study instructions added in response to performance funding are highlighted in appendix F, "Guidelines for Self-Study Document." In short, the established mechanism for program review at UTK is sufficiently comprehensive that it can be adjusted to provide data in response to the criteria established for funding. In fact, an already strong review mechanism has been improved by the systematic collection of data on student outcomes and by comparison of these data with departmental objectives for student performance.

The need to furnish data for performance funding has provided the opportunity to strengthen program review at UTK. Now, departments must not only gather objective data concerning program effectiveness, but also define a common core of information to be gathered. This ensures comparability among self-studies and reports, making them more useful for institutional planning, budgeting, and decisionmaking. Furthermore, the THEC guidelines for performance funding require that information be gathered about program quality on a five-year cycle. This has necessitated some tightening and revision of the schedule for internal review at UTK, and it has added an increased risk of superficiality. The need to meet the THEC criteria, however, also encourages advanced planning, such as the early scheduling of an entire multiyear cycle. Thus, departments have the opportunity to collect better and more extensive data to meet both internal and external requirements. Advance scheduling of program reviews also ensures that the most qualified consultants are available when needed.

The addition of objective data to the self-study has enriched the quality and volume of information available for assessing program strengths and weaknesses, and it has also increased the ability of central administrators to use review documents for planning and budgeting. Furthermore, the study provides a list of specific actions required for implementation. Following submission of reports from internal and external reviewers, the department head has one month to respond in writing to the central administration. Then the college dean organizes and directs the

follow-up sessions, which include departmental representatives and key central administrators. In these sessions, each recommendation and response is reviewed, and a list of appropriate actions is developed and forwarded to those administrators responsible for budget review and strategic planning.

The current concern at UTK is to persuade the THEC to give greater weight to program review as a means of program assessment, as opposed to assigning the majority of credit under Variable II to objective testing of students in the major field. While objective measures have an important place in the review process, they are too limited in scope to provide clear directions for change of the department's or institution's goals. The self-study portion of the review process, for example, has significant benefits by itself, including the impetus to examine policies, practices, procedures, and records. Although important, such work is often neglected until a crisis occurs. According to Kells (1983, p. 17), a self-study "should contain an informed attempt to clarify organizational goals and to assess achievement of the goals for purposes of improvement." Going through the self-study phase is extremely beneficial for faculty members and administrators, and the resulting document provides a useful basis for peer review and evaluation. Testing for student outcomes does not by itself provide a similarly comprehensive view of a program's strengths and weaknesses.

## Conclusion and Recommendations

Whether or not an institution is asked to respond to external evaluation criteria, it can benefit enormously from an internal program review that is comprehensive, emphasizes outcomes and peer review, and feeds into planning and resource allocation. The self-study should include statements of program functions and objectives, descriptions of the program, its faculty, and students, an assessment of library resources and physical facilities, information about student outcomes, and other objective evidence of program quality.

Program review may seem expensive, but its costs should be viewed as investments in the cultivation of excellence at the institution. In addition to the time and materials involved in the self-study, costs include the expenses and honoraria paid to external reviewers, as well as the expenses incurred by faculty members and administrators who participate as internal reviewers. Obviously, much time must be devoted to studying the program and considering recommendations for improvement, both of which are essential to increased quality and efficiency. But, when viewed from the perspective of potential results, program review is clearly cost-effective.

Although program review is complex, the process itself can be evaluated for effectiveness in a number of ways. First, program review can be tested for soundness. What are the components and how do they interrelate? What is the quality of the self-study? Do the credentials of the reviewers make their recommendations credible? Second, the role of

program review in the institutional planning process can be examined to see how recommendations are addressed and priorities determined. Third, the schedule of reviews and follow-up can be examined to see if the process is carried out consistently and conscientiously. Fourth, the role of outcome measurements in the process can be examined to see if deficiencies are addressed over time. Departments and colleges can indicate the steps that they have taken to improve student outcomes with respect to instructional objectives. Some departments can even use national reputational surveys to demonstrate improvement in program effectiveness. Finally, the central administration can provide detailed information about the recommendations that were acted upon and the types of changes that were made. Actions designed to effect improvements can take many forms, including reallocation of internal resources to provide for equipment needs or additional faculty; adjustments in curriculum patterns, course structure, or instructional methods; changes in ways to advise students; and fine-tuning of the instruments used to assess student outcomes. For these reasons, state agencies need not avoid using reviews that are designed by particular institutions for the purpose of assessing program quality across different institutions.

In sum, a well-designed and implemented review process of academic programs subsumes and integrates the major criteria used in performance funding. Indeed, this form of program evaluation represents the summation of all the assessment processes in the THEC program, and as such, it should be considered the core element of the Instructional Evaluation Schedule. Good review processes include assessments of program outcomes as well as program goals and the adequacy of invested resources. A detailed plan for program review should contain a variety of means for measuring the achievement of program objectives by students in the major field, including testing, performance ratings of students or graduates, external evaluation of theses and dissertations, placement of graduates in related jobs or in accredited graduate programs, and outstanding accomplishments of graduates. Good processes also allow the use of informed, expert judgment in the assessment of program quality. The essence of performance funding is contained in the link it establishes between explicit dimensions of effectiveness and the process of resource allocation. Program review mirrors this process at the program or department level, but only if review results are constantly and conscientiously linked to budgeting and program improvement.

If a review process meets all of the above criteria, it will eventually become the most effective tool available for use in performance funding. When a good process is in place the actual improvements, rather than the measurement process per se, can be used to judge program quality and effectiveness—the ultimate goals of both the THEC and the participating institutions throughout the state of Tennessee.

## References

Barak, Robert J. *Program Review in Higher Education. Within and Without.* Boulder, Colo.: National Center for Higher Education Management Systems [NCHEMS], 1982.

House, Ernest R. "Alternative Evaluation Strategies in Higher Education." In *Designing Academic Program Reviews,* pp. 5-15. New Directions for Higher Education, no. 10. Edited by Richard F. Wilson. San Francisco: Jossey-Bass, 1982.

Kells, H. R. *Self-Study Process.* 2nd ed. New York: Macmillan, 1983.

Wallhaus, Robert A. "Process Issues in State-Level Program Reviews." In *Designing Academic Program Reviews,* pp. 75-87. New Directions for Higher Education, no. 10. Edited by Richard F. Wilson. San Francisco: Jossey-Bass, 1982.

PART THREE

# An Assessment of Performance Funding



103

CHAPTER NINE

# Performance Funding and Institutional Response: Lessons from the Tennessee Experience

*By Peter Ewell*

MANY STATES HAVE recently changed their higher-education policies in response to rising demands for quality in publicly funded colleges and universities. Historically, such calls for improvement are not new. Indeed, American colleges and universities have undergone repeated cycles of self-assessment and improvement, and they have renewed their commitment to quality through changes such as curriculum reform, instructional delivery, and student recruitment and selectivity. Today, however, colleges and universities must respond to external demands for change, and they must do so with extremely limited resources. The challenge is to do more with less or face the loss of revenues previously taken for granted. Rising demands for demonstrable improvement must be met without either adding new programs or increasing overhead costs associated with additional regulation and accountability.

Tennessee's peformance funding initiative, in place now for five years, represents a comprehensive attempt to meet this challenge. Many states have initiated programs, regulations, or funding mechanisms designed to accomplish a similar purpose, but no other state system of higher education emulates completely the Tennessee approach. Because the Tennessee experiment embodies many elements in common with programs in other states, however, its lessons contribute to an understanding of the process involved in such state level initiatives for improvement of the quality of higher education.

The response of one major institution of higher education in Tennessee, the University of Tennessee, Knoxville (UTK), to the challenge of performance funding can tell us a lot about the process of establishing a program to assess outcomes at a major university. The choices to be made about what to measure, how to measure, and above all, how to disseminate and use the results of measurement are illustrative of the dilemmas faced by all institutions of higher education embarking on this road. Recently, the

number of colleges and universities attempting to implement such campus wide programs has increased. Some have done so solely in response to pressure from external governing or accrediting bodies. Many more have done so because they recognize the dividends that such programs can pay in terms of external credibility, internal planning and budgeting, and the ability to attract and retain better students. The UTK program, like performance funding itself, reflects a distinctive institutional environment. It merits close attention both because it is unique and because it has much in common with the self assessment efforts of other institutions across the country.

## Performance Funding in a National Context

In the year since NIE issued *Involvement in Learning* (1984), several states have launched programs aimed at improving the effectiveness of undergraduate instruction. Many of these programs share features with Tennessee's initiative in performance funding. This is partly because other states have learned from Tennessee's experience and have incorporated some elements of Tennessee's program into their own. It is also partly because the general structure of higher education and the tools that state governments have available to address higher education issues are similar across states.

To effect qualitative improvements in higher education at the institutional level, state governments basically have two tools with which to work (Jones 1982, 1984, Ewell 1985). First, they can change the budgetary-allocation process so that it provides special funds for needed improvements or rewards and encourages successful performance. In many states, this will mean changes in established funding formulas that govern the allocation of dollars to individual institutions. In others, modifying allocation may involve setting aside resources to be distributed by means of a separate process. Second, states can exert direct control over institutional behavior through regulation or statute. In some cases, this will mean establishing requirements for specific performance on the part of currently enrolled students or graduates. In others, it may mean either specifying that existing resources be expended in designated ways or requiring that institutions adopt explicit evaluation activities or program reviews. Other external agencies have joined state governments in using the second approach. Prominent among them have been regional and professional accreditation bodies. The former increasingly require that institutions demonstrate that certain processes are in place. The latter increasingly require evidence of demonstrated performance on the part of program graduates.

One of the interesting features of Tennessee's performance funding initiative is that it makes use of both these mechanisms. Because it affects actual dollar allocations, the program resembles purely financial approaches that manipulate the behavioral incentives for institutions by changing the parameters of the established allocation formula or process. But because it

goes beyond such manipulations to specify explicit performance criteria and reward particular types of activities, the Tennessee program has much in common with the second approach. To date, Tennessee's performance funding process constitutes the only such intersection, but many states have implemented programs along one dimension or the other.

Probably the most powerful single lever that state governments can use to influence institutional behavior is the method they choose to allocate resources. But the effectiveness of the allocational lever tends to be limited. Given substantial existing investments and the need to maintain the asset, shifts in actual dollar allocations among institutions tend to be long-term and marginal. Recent reviews of state-level allocation mechanisms indicate that approximately half of the states currently allocate funds to institutions on the basis of an established funding formula (Brinkman 1984). Although a widespread phenomenon, formulas are far from standard. They vary widely from state to state and over time within each state. Because most are enrollment-driven, however, established formulas have recently come under a great deal of criticism. Designed to accommodate institutional growth in the sixties and early seventies, colleges and universities often find themselves ill served by such mechanisms when enrollment is down. As Brinkman notes,

> Formula budgets were one of the mechanisms developed to deal with enrollment growth, and they were quite effective. Had the originators realized fully the extent to which many of the formulas fostered growth and not merely accommodated it, they might have been a little more careful. ... The original growth incentives threaten now to result in lower admission standards and questionable recruiting practices. [1984, p. 28]

Other criticisms of formulas have arisen in the context of increasing public demands for qualitative improvements. First, formulas are sometimes uniformly applied across quite different kinds of institutions. The result is that, as the incentives that formulas provide for similar behavior become clear, they tend to have a leveling effect on institutions (Gross 1973). Second, formulas concentrate on input measures and measures of activity level that are readily quantifiable (for example, enrollments, faculty, and gross square footage). As a result, they are not related to the goal of higher education, namely improvement in student learning and development (Pickens 1982). Acting in concert, these two drawbacks tend to reward quantitative growth over qualitative improvement.

Due to these drawbacks, many states have been experimenting with ways to mitigate the negative effects of formula budgets while preserving at least some of the simplicity, clarity, and impartiality of the formula approach. Among the trends cited by Brinkman (1984), for example, are the tendencies to (1) create "buffering mechanisms" that mitigate the impact of enrollment shifts through multiyear averaging or by establishing institutional base guarantees, (2) include other kinds of measures of activity level

(for example, student headcount) in formulas traditionally based solely upon student credit hours taught by level and discipline, (3) make formulas more complex in order to take into account differences in cost and production functions for different parts of the institution, and (4) improve the cost calculations themselves. A general tendency, as reported by a recent national survey of state funding agencies, has been progressively to decouple the traditionally linear relationship between enrollment and allocation (Leslie 1983).

At the same time, concern for qualitative improvement has induced states to modify the allocation process in additional ways. In most cases, this involves setting aside funds for identified purposes. The most direct use of such funds is the one embodied in Tennessee's approach: use of incentive funds to "buy results" at the institutional level. The Tennessee Higher Education Commission (THEC) program is almost unique in this respect, but some parallels exist in Florida's challenge grant program for funding new endowed chairs. Parallels also exist with New York's established degree capitation grant program (Bundy Funds) for private institutions. This grant program provides institutions with specific funds for each degree successfully produced.

Most state-level improvement programs, in contrast, are even more special-purpose and are put in place to foster a specific qualitative improvement, type of program, or institutional practice. Tennessee's own "Centers of Excellence" program provides an example of such an approach. States such as Virginia have used set-aside funds to establish grantlike programs that foster innovations in curriculum or administration at the institutional level. New Jersey has recently launched a multifaceted program with similar features. Finally, states like Missouri have experimented with allocating funds for program improvement on the basis of projected "returns on investment" in actual outcomes (McClain 1984).

As several observers have noted recently (for example, Brinkman 1984; Spence and Weathersby 1981), such "set-aside" and categorical grant programs have the major virtue of letting enrollment-driven allocation formulas do what they do best, that is, provide base funding in appropriate amounts to maintain basic operations and assets while providing additional, targeted incentives for carefully selected qualitative improvement areas. And the most likely future trend is further differentiation between these two functions and funding approaches.

What are the implications of these trends for programs such as the one in Tennessee? First, the THEC approach is solidly within the national trend in two senses. It operates at the margin (a 5 percent set-aside is certainly substantial, but by no means revolutionizes the total funding picture), and it explicitly distinguishes funds that are designed to induce change from those designed to maintain business as usual. Furthermore, the THEC approach has managed to preserve within the qualitative component many of the virtues of a formula approach. By setting explicit criteria

for performance as the basis for allocation, the THEC program helps depoliticize and regularize the allocation process. All institutions are subject to performance evaluation on the basis of the same criteria, and the public nature of the process makes it likely that such criteria will be uniformly applied. From a purely structural perspective, this is a powerful advantage. But precisely because of this advantage, many states may be reluctant to take this approach. Explicit performance criteria, when spelled out in advance, can bind both states and institutions to a set of mutual obligations. Institutions know that if they perform according to criteria, they will receive benefits accordingly; but providers also know that if performance is forthcoming, they will have to deliver. Given the political nature of the allocation process, many states simply will not allow their discretion to be so constrained.

Tennessee's initiative in performance funding is also illustrative of a second major trend that responds to demands for quality improvement: the need for explicit measures of institutional and program performance. Rather than coupling performance incentives with the allocation process, a number of states, as part of their exercise of regulatory authority, have opted to require certain kinds of activities.

A few states have responded to public concern about academic quality by mandating that students demonstrate specific levels of performance. "Rising junior" testing programs are now in place in Florida and Georgia, and are being actively explored by several other states. In Florida, for example, students in public institutions completing their sophomore year must successfully pass a statewide College Level Academic Skills Test (CLAST) as a condition for enrollment with junior standing. In Mississippi and Missouri, graduates of teacher education programs must score above designated levels on a standarized achievement test as a condition of graduation. Finally, in South Dakota, the public higher-education system has implemented a requirement that all students be tested by an established standardized examination in their major field before they can graduate.

Even where explicit requirements for student performance have not been put in place, states have used their regulatory authority to induce qualitative improvement. Often this means requiring that institutions provide evidence of program effectiveness through assessment and testing. For example, an element of New Jersey's broad-based, quality-improvement initiative involves statewide assessment of student learning in general education for the purpose of curriculum improvement. The state is currently working with testing agencies to design the necessary instruments that will be administered next year. Colorado has in place a somewhat similar assessment requirement for implementation by all public institutions by 1989. Both these examples have much in common with Tennessee's program in that measurement results are used in the aggregate to provide evidence of program strengths and weaknesses rather than to decide the fate of individual students.

Additionally, a recent trend has been for legislative and executive audits of public colleges and universities to be conducted as performance audits. Traditionally, audits have been almost exclusively confined to financial regulation, ensuring that institutions spend monies according to plan. A second mechanism is provided by institutional and statewide processes of program review. As indicated by a recent survey of state higher-education agencies, more than two-thirds of the states now have in place explicit procedures for academic program reviews, and most of these are of recent origin (Barak 1982). Another trend is for such reviews to be increasingly oriented toward performance and require that the program provide data about student achievement and subsequent educational and job success.

In its identification of explicit criteria for assessing institutional performance, the performance funding program of the THEC is consistent with these trends. In setting these criteria broadly, however, and in allowing institutions considerable latitude in how the criteria are to be met on each campus, the Tennessee approach most resembles recent trends in accreditation. This is a particularly healthy development as it recognizes, first, that there are legitimate differences among institutions in a state system and, second, that uniform performance may be an inappropriate basis for assessment. Maintenance of such an approach, however, depends upon the existence of a great deal of trust between the institution and the funding or regulating authority. And, across the nation, such a level of trust is currently more the exception than the rule.

## Dimensions of Institutional Performance

One major problem of any performance funding or assessment mechanism is the difficulty of defining the concept "institutional performance." Regulatory approaches such as those discussed above have included definitions, albeit narrow ones, funding mechanisms as a rule have not. There are, of course, many reasons for this. One major reason is simply that funders do not want to reduce their discretion in seeking quality. A reason for clearly separating funds for quality improvement from general formula-driven support (the approach that many states have taken) is to allow funds of the former kind to be flexibly allocated as perceived needs and opportunities arise. Clear but narrow definitions of performance, like the one utilized in Tennessee, are seen by many to reduce such flexibility.

But at least as important an obstacle is the sheer difficulty of defining the concept "effectiveness of instruction." The first problem is to determine the appropriate level of analysis for assessing effectiveness. Most observers have concentrated on assessing effectiveness at the institutional level, and indeed this makes a great deal of sense. It is quite possible, however, to take a different view of effectiveness. From this perspective, efforts to improve quality are viewed against a template of statewide system effectiveness that holds that institutions may vary greatly in performance and mission so long as the system as a whole is meeting statewide educational needs

effectively (Jones 1984). For example, a statewide effectiveness approach may take into account outcomes regarding economic development, access to and distribution of the benefits of higher education to different segments of the population, and the cost-effectiveness of alternative investments in higher education (Ewell 1985). Finally, approaches such as Florida's concentrate on performance at the individual level. The CLAST examination is a common hurdle, but one faced by individual students rather than institutions or programs.

Even when the institutional level is clearly chosen as the analytical focus, definitional difficulties abound. In most cases, the concept "performance" is defined in terms of instructional effectiveness. What exactly does this mean? Many dimensions of educational outcomes are available against which to assess actual performance (Lenning, Lee, Micek, and Service 1977, Bowen 1977). No performance evaluation mechanism can assess them all. Choosing which outcomes to assess and how to assess them is thus a major challenge when constructing such a program.

Most observers have addressed the assessment of higher educational outcomes in terms of one of the following four basic dimensions (Ewell 1984):

- *Knowledge outcomes,* which refer to changes in cognitive learning. Knowledge outcomes can refer to either general knowledge (for example, knowledge of Western civilization) or knowledge of specific academic fields or disciplines.
- *Skills outcomes,* which refer to changes in student abilities to perform particular functions or tasks. Like knowledge outcomes, these can refer to either general skills (for example, reading comprehension, problem solving, or leadership) or task- or occupation-specific skills.
- *Attitude/value outcomes,* which refer to affective changes occurring as a result of the college experience, for example, tolerance for diversity, attitudes toward self, and satisfactions with the educational environment.
- *Behavioral* manifestations of each of the above, as revealed by factors such as student employment and job performance/success after graduation, subsequent educational experiences and performance, and social and familial behavior.

Each of these major dimensions contains a host of subcategories, and an important issue when defining the concept "performance" is to decide which dimensions should be chosen. Traditional short-term evaluative measures in elementary and secondary education, for example, have concentrated almost exclusively on the cognitive dimension. Many classic assessments of postsecondary education, in contrast, have dealt primarily with affective development (for example, Astin 1977, Chickering 1969) or concentrated on the behavioral dimension of impact after graduation (for example, Bowen 1977). Most of the recent moves toward performance

evaluation at the state level follow the elementary/secondary tradition. Initiatives such as those taken in Florida, Missouri, Mississippi, and South Dakota, for example, are all heavily dependent upon actual cognitive assessment of student learning.

The THEC performance funding approach is somewhat different. Certainly, cognitive development is a strong part of the THEC specification of institutional performance, as evidenced by the fact that two criteria deal with cognitive knowledge of general topics and specific fields. The ACT COMP instrument used to assess general knowledge, however, also involves a substantial assessment of component academic skills. In its long form, students are asked to write, speak, and organize knowledge in response to different kinds of stimuli. Even the short form of the ACT COMP emphasizes reading and problem-solving skills. Additionally, the THEC program contains a criterion of student satisfaction, although it is weighted relatively lightly compared to other dimensions. Finally, the criteria of program accreditation and program review permit several other basic dimensions to be assessed indirectly. Many accreditation processes include assessments of actual student performance or certification in the profession after graduation. Furthermore, many processes that assess academic programs allow for or require similar evidence of actual student success.

Because of its wide range of potential coverage, the set of THEC criteria allows each institution to demonstrate its particular strengths. More important, with the exception of the criterion of general knowledge, the THEC criteria offer enough latitude in interpretation so that it is up to each institution to determine exactly how it will demonstrate each type of performance. Placing the responsibility for the way performance assessment is to be implemented has both costs and benefits. On one hand, different institutions can (and do) choose somewhat different ways of approaching the problem, depending upon what they perceive to be their strengths. Indeed, several of the criteria, including the important general knowledge criterion, state alternatives for earning points. This kind of flexibility is reinforced by the program review criterion, a criterion that provides points for the existence of a process, whatever the actual outcome.

On the other hand, institutional choice with respect to the mechanics of measurement means that there is considerable potential for abuse of the system if the level of trust between the state agency and the institution breaks down. By their very nature, statistics about performance outcomes are difficult to interpret and verify. Measurement error is inherent in such assessment procedures. And measurement error is potentially exacerbated when data are gathered for small, locally selected samples rather than for the entire population at each institution.

Most institutional and state assessment programs include some coverage of all four basic outcomes dimensions noted above. But, in most cases, the choice of instruments used and the relative emphasis given to each dimension in the assessment process vary considerably depending

upon the institutional mission and the intent of the assessment. As the UTK experience shows, different types of data have very different impacts on the decisionmaking process. The value of the data collected for purposes set by the THEC varied considerably across the campus and, consequently, were given quite different emphases.

### The UTK Response to Performance Funding

Initiatives such as performance funding are effective only if individual institutions actually make changes to improve programs and services. The UTK response to the THEC initiative is an excellent example of the way such a process ideally works. But it is important to stress that considerable planning, investment, and sensitivity has been required to make it work. Indeed, careful examination of the manner in which the UTK program has been developed tells us a great deal about what constitutes success in an institution's self-assessment effort.

Since the publication of the NIE report *Involvement in Learning* (1984) one year ago, many colleges and universities have embarked upon testing and measurement as a part of undergraduate instructional improvement. Their motives are many. Some local assessment programs arise out of a new curriculum or learning philosophy. For example, the well-known testing/assessment program at Alverno College was a direct result of a major curriculum realignment over a decade ago. All students at Alverno are regularly assessed by a variety of mechanisms that are both independent of the regular course grading process and include multiple assessments of skill as well as in-depth personal interviews. At the same time, Alverno has undertaken numerous studies to validate externally its cross-disciplinary curriculum (Mentkowski and Doherty 1984). Emerging assessment programs at King's College and Southwest Texas State University have similar roots. Both arose out of a desire to evaluate the effectiveness of extensive revisions in undergraduate general education.

Other institutions have embarked upon self-assessment in order to demonstrate that their graduates are nationally competitive. For example, since 1975, Northeast Missouri State University has conducted a wide range of assessment studies about the value added to students' learning over time. The majority of these studies rely on nationally normed test instruments such as the American College Testing Program College Outcomes Measures Project (ACT COMP), the ACT Assessment (administered in a test/retest format), Graduate Record Exam (GRE) Field Examinations, and a variety of standardized professional and preprofessional examinations. Applying the results of these assessments across the curriculum has led to both gains in achievement and increases in the number of high-quality students attracted to the university (McClain and Krueger 1985). Indeed, the recruitment impact of assessment information is increasingly being recognized by colleges and universities. For example, Montana State University recently used data on student success to develop a targeted

recruitment program aimed at sophomore and junior level high school students. What have been the effects of such programs? Early evidence suggests that one of their most important functions is to stimulate faculty discussion about and involvement in the curriculum (Ewell 1985). But such effects do not follow automatically when an institution initiates an assessment program. Indeed, as the UTK experience shows, the process of building faculty and administrative involvement can be both slow and painful. Efforts have been successful partly because they have fostered widespread participation from the outset and allowed sufficient time for an effective assessment process to be developed. They have also been successful because attention is paid constantly to the relationship between the assessment information and the identification and solution of particular problems faced by faculty members and administrators.

Often the key to effective use of instructional evaluation information is to break down results by curriculum, skill area, or student behavioral group. For example, at Northeast Missouri State University, evidence from assessments of value added in general education showed that students were not retaining mathematical skills in their first two years of college. These results were used to make changes in the mathematical content of a number of general-education courses (McClain and Krueger 1985). Results can also lead to changes in the sequencing or structure of individual curricula. At Mt. Hood Community College, for example, changes were made in the course sequence in several occupational programs based on evidence of student performance on the job (Stevenson, Walleri, and Japely 1985). Finally, when results are available for different types of students, student satisfaction surveys can be used effectively to improve student services and structure student-retention programs. At North Carolina State University, for example, data from such surveys were used to help build and evaluate a set of academic-skills programs with components explicitly targeted for different types of students.

Despite growing instances of success, however, comprehensive assessment programs have until recently been a relative rarity among institutions of higher education. For the most part, this is because few immediate incentives exist for individual institutions to adopt them. An important impact of statewide programs is that they create immediate incentives for constructing programs such as performance funding, which as they develop over time gradually reveal their considerable internal benefits. In many ways, this was the course of the process at UTK. Clearly, in the absence of performance funding, UTK would not have developed the program it did. But the evolution of instructional evaluation at UTK also reveals that considerable care was taken by campus administrators to ensure that the assessment program both generated useful results and remained true to institutional purposes.

Indeed, the UTK experience confirms many lessons that are now emerging as a result of successfully establishing comprehensive assessment programs. Foremost among these lessons are the following:

- Instructional evaluation requires strong administrative commitment and an effective administrative center. As Trudy Banta describes, a key aspect of the UTK program was the assignment of instructional evaluation to the Learning Research Center, an established and visible office at the university. This assignment allowed many types of data-gathering activities to be effectively coordinated. Furthermore, adequate initial investment in the center allowed measurement professionals to provide technical assistance to departments and units in the design of tests and surveys and in the interpretation and use of assessment results. Because of this, UTK has made much more effective use of information than have campuses in which programs lack these features.
- A successful program collects data from multiple sources using multiple methods. Following the framework of the THEC guidelines, the UTK program included three major types of data gathering: ACT COMP testing to assess student gain in general education, major field testing, and administration of a student satisfaction survey. While each of these investigations has proven useful in its own right, the combination provides a particularly strong foundation for instructional evaluation. No currently available test or instrument is sufficiently valid to provide an exclusive basis for curricular or policy action. Multiple measurement not only provides a check against drawing invalid conclusions but also allows much more complex phenonmena to be investigated. The UTK policy of providing to each department a packet that combines the results of many studies has proven particularly useful. Other campuses have strongly confirmed the utility of this procedure.
- Successful assessment requires a good match between what is measured and what is taught. While the THEC guidelines require administration of the ACT COMP exam to assess student gain in general education, the choice of instruments and methods for assessing other dimensions of student outcomes is left to the individual campus. As illustrated by chapters 3 and 6, at UTK the choice was far from easy. In the case of student satisfaction, careful review of existing published instruments revealed that none adequately tapped important elements of the campus environment. Moreover, for many departments, standardized instruments such as the GRE were either not available or were not suitable for assessing student achievement in particular major fields as taught at UTK. As a result UTK has invested heavily in instrument development, but this investment has paid off well because the design of assessment instruments closely matches actual curricular content and objectives.
- The most fruitful assessment programs begin with full involvement of faculty members in the initial design phase, and end with faculty members as active participants in the interpretion and use of the results. The UTK use of faculty/administrative task forces to choose

instruments and develop assessment policy provides an excellent model of how an instructional evaluation program ideally begins. The task forces were composed of faculty members from different parts of the institution, and they were supported by considerable technical expertise in testing and measurement. The task force members, in addition to providing an initial design for the program, gave UTK a nucleus of committed faculty members who could help implement the program in its crucial first years. For example, members of the task force on student satisfaction helped design the instrument that was ultimately used, and some members of both this task force and the task force on student achievement in the major field provided their own departments as pilots for first-year implementation.

- Instructional evaluation is ineffective if there is no formal linkage between assessment and planning and budgeting. The THEC program includes a reward for such linkages through its program review criterion, and as Homer Fisher notes in chapter 2, the primary use of assessment data at UTK has been in the program review process. As Fisher describes, the best use of such data is in conjunction with other more traditional indicators of program performance, including available resources and their quality, enrollment/activity levels, and costs. At the same time, the UTK program encourages local use of the data by deans, department heads, and unit managers themselves. By making results directly available to these mid-level decisionmakers and providing the necessary technical help to interpret the data and draw appropriate policy conclusions, the UTK program contains multiple avenues for ensuring that instructional evaluation information, once collected, is effectively and appropriately used.
- Successful assessment programs take time. As Robert Levy describes in chapter 1, the THEC performance funding program has been some 10 years in the making, and it has evolved considerably from the earliest proposals into a flexible, workable program. The UTK experience parallels this process at the campus level. An initial planning year in which task forces met to consider the shape and instrumentation of the program was an important ingredient of success. Similarly, the decision to implement assessment the following year on a pilot basis and use available grant funds to support several interested departments in their efforts to gather and use information yielded considerable dividends. Based on the results of pilot projects, other departments were able to see the benefits that resulted from having gathered evaluation data and were less reluctant to undertake efforts of their own. Most important, a realistic schedule allowed development and evolution of a program that reflected departmental distinctions and considerable faculty input. It is important to stress that assessment programs such as those at Alverno and Northeast

Missouri State University also evolved over time and required considerable initial planning. Institutions contemplating such programs should consider similar paths of development and not be too hasty in expecting immediate and dramatic results.

The UTK instructional evaluation program was substantially shaped by the statewide performance funding program. As a result, many of the features of the UTK instructional evaluation program were structurally determined from the outset by external influences. But the way UTK administration proceeded within this external structure is exemplary. Regardless of statewide action, any campus contemplating an assessment program would do well to examine the UTK experience.

## A Look at the Future

What is the future of statewide quality incentive programs such as the THEC performance funding program? Recent events suggest that many more states will in the next few years join Tennessee in launching initiatives aimed at improving undergraduate instructional quality. Few such emerging programs, however, contain the kinds of direct linkages between dollars and performance as the Tennessee experiment. Based upon these trends, it seems unlikely that performance funding mechanisms will ever replace enrollment-driven formulas or size-based criteria for allocating the bulk of a state's higher education dollars. At the same time, use of a performance funding mechanism for distributing quality improvement funds at the margin, once such funds have been explicitly identified, is quite attractive. And it is important to note that the actual sums involved for particular institutions can be quite sizable. For example, in FY85, UTK earned a total of $3.5 million in benefits from the performance funding program.

The theoretical and practical advantages of an explicit performance funding program for distributing such resources are many. Like a formula, the criteria embedded in the process are uniformly applied, and they may tend to depoliticize the allocation process. And, like regulatory approaches, the criteria may be adjusted to emphasize particular areas of performance deemed to be of statewide interest or importance. Furthermore, the kinds of criteria chosen may emphasize either good practice or explicit performance. The original THEC program, for example, did not mandate particular outcomes or instruments in any of its criteria, while the current version does both. Most important, a performance funding approach provides positive, immediate incentives for an institution to conduct appropriate and ongoing internal assessments of its programs and services. As the previous chapters have testified, the real benefits of performance funding are seen in the curriculum modifications and improvements made by institutions such as UTK.

At the same time, the conditions under which a successful performance funding program can be implemented may well be rare. By its very

nature, such a system entails an unusual level of trust between the institution and the governing agency. Agencies, and ultimately legislatures, must be willing to state once and for all what they are looking for and reduce their own discretion in rewarding excellence. In many political situations, this may be a lot to ask. Similarly, institutions must accept such initiatives openly and conduct assessments in as systematic, valid, and participatory a manner as possible. Several institutions in many different states have already seen the abuses of inappropriate comparison in endeavors such as statewide cost and load studies. In outcomes studies, the potential for abuse is equally great. In the absence of a high level of trust, sharing institutional performance data constitutes for any college or university an act of considerable courage.

On balance, however, the future of such initiatives is good. Although states will continue to vary in their approaches to improvement of quailty, it seems clear that an increasing number will begin to address the issue in some fashion. As they do so, the advantages of the performance funding approach will probably become clearer. Moreover, no matter what the ultimate outcome, the choice for many states of which mechanism to use will be partly based upon the experience of other states.

The same is true for individual institutions that are contemplating construction of a comprehensive program of student assessment. Performance funding can be an institutional as well as a state intitiative, and many individual colleges and universities are currently experimenting with campus-based programs that assess outcomes and link resources with instructional evaluation. Because the payoffs of efforts in areas such as student recruitment, retention, and improved instructional quality are becoming increasingly known, the future of campus-based assessment seems bright regardless of what states ultimately decide to do. For these reasons, Tennessee's experience with performance funding and the institutional responses that it has induced provides an excellent base on which to build.

## References

Astin, Alexander W. *Four Critical Years. Effects of Colleges on Beliefs, Attitudes, and Knowledge.* San Francisco: Jossey-Bass, 1977.

Barak, Robert J. *Program Review in Higher Education. Within and Without.* Boulder, Colo.: National Center for Higher Education Management Systems [NCHEMS], 1982.

Bowen, Howard K. *Investment in Learning. The Individual and Social Value of American Higher Education.* San Francisco: Jossey-Bass, 1977.

Brinkman, Paul. "Formula Budgeting. The Fourth Decade." In *Responding to New Realities and Funding,* pp. 21-44. New Directions for Institutional Research, no. 43. Edited by L. Leslie. San Francisco: Jossey-Bass, 1984.

Chickering, Arthur W. *Education and Identity.* San Francisco. Jossey-Bass, 1969.

Ewell, Peter T. "Some Implications for Practice." In *Assessing Educational Outcomes,* pp. 111-20. New Directions for Institutional Research, no. 47. Edited by Peter T. Ewell. San Francisco: Jossey-Bass, 1985.

———. *The Self-Regarding Institution. Information for Excellence.* Boulder, Colo.: NCHEMS, 1984.

Gross, Francis M. *A Comparative Analysis of the Existing Budget Formulas Used for Justifying Budget Requests or Allocating Funds for the Operating Expenses of State Supported Colleges and Universities: A Dissertation Summary.* Knoxville: The University of Tennessee, 1973.

Jones, Dennis P. *Higher-Education Budgeting at the State Level. Concepts and Principles.* Boulder, Colo.: NCHEMS, 1984.

———. *Data and Information for Executive Decisions in Higher Education.* Boulder, Colo.: NCHEMS, 1982.

Lenning, Oscar T.; Lee, Yong S.; Micek, Sidney S.; and Service, Allan L. *A Structure for the Outcomes of Postsecondary Education.* Boulder, Colo.: NCHEMS, 1977.

Leslie, Larry L. "Recent Financing Developments in the Fifty States." In *Survival in the 1980s: Quality, Mission, and Financing Options,* pp. 185-92. Edited by Robert A. Wilson. Tucson: Center for the Study of Higher Education, University of Arizona, 1983.

McClain, Charles J. *Degrees of Integrity: A Value-Added Approach with Undergraduate Assessment.* Washington, D.C.: American Association of State Colleges and Universities [AASCU], 1984.

McClain, Charles J., and Krueger, Darrell W. "Using Outcomes Assessment: A Case Study in Institutional Change." In *Assessing Educatonal Outcomes,* pp. 33-46. New Directions for Institutional Research, no. 47. Edited by Peter T. Ewell. San Francisco: Jossey-Bass, 1985.

Mentkowski, Marcia, and Doherty, Austin. *Careering After College. Establishing the Validity of the Abilities Learned in College for Later Careering and Professional Performance.* Milwaukee, Wis.. Alverno College 1984.

National Institute of Education. *Involvement in Learning. Realizing the Potential of American Higher Education.* Report of the Study Group on the Conditions of Excellence in American Higher Education. Washington, D.C.: Government Printing Office, 1984.

Pickens, W. H. "What's Ahead for Higher Education?" *Journal of the National Association of College Auxiliary Services,* April 1982, pp. 8-12.

Spence, David S., and Weathersby, George B. "Changing Patterns of State Funding." In *Challenges of Retrenchment. Strategies for Consolidating Programs, Cutting Costs, and Reallocating Resources,* pp. 226-42. Edited by James R. Mingle. San Francisco: Jossey-Bass, 1981.

Stevenson, Mike; Walleri, R. Dan; and Japely, Saundra M. "Designing Follow-Up Studies of Graduates and Former Students." In *Assessing Educational Outcomes,* pp. 81-91. New Directions for Institutional Research, no. 47. Edited by Peter T. Ewell. San Francisco: Jossey-Bass, 1985.

# Appendixes

APPENDIX A

# THEC Standards of Performance

## 1979

### Performance Variable I—Program Accreditation

| Standards of Performance | Suggested Point Allocation |
|---|---|
| An institution will be awarded points on this variable based on the percentage of eligible* programs accredited. For example, an institution having 30% of its eligible programs accredited would be awarded 6 points. | 20 |

* A program is defined as "eligible" if there is a COPA-approved agency or organization which accredits programs for that field and degree level—and especially if a Tennessee institution holds accreditation for that particular degree and field.

TABLE *(continued)*

**Performance Variable II—Performance of Graduates on a Measure of Specialized or Major Field Competence**

| Standards of Performance | Suggested Point Allocation |
|---|---|
| (1) Institution has assessed performance* of a representative sampling of graduates** in one or more of its major program fields within last three years. | 5 |
| (2) Institution has assessed performance of a representative sampling of graduates in the majority of program fields during the past three years. | 10 |
| (3) Institution has assessed performance of a representative sampling of graduates in majority of its program fields within the past three years and can demonstrate that the performance of its graduates ranks with or above the performance of graduates from similar institutions in majority of these fields. | 12-20 |

* The measure of performance must be an assessment instrument/procedure constructed external to the institution—with normative standards available for state, regional, or national samples. Examples would include the GRE field tests by ETS, state or national licensing examinations, professional field tests such as the National Teacher exams, etc.

** The number of graduates assessed must be sufficiently representative to permit statistically sound inferences to all graduates in that field.

TABLE *(continued)*

**Performance Variable III—Graduate Performance on a Measure of General Education Outcomes**

| Standards of Performance | Suggested Point Allocation |
|---|---|
| (1) The institution has assessed the performance of a representative sampling of graduates* on a measure of general education outcomes** on a pilot or one-time basis during last three years. | 5 |
| (2) The institution has an ongoing program to assess the performance of its graduates on a measure of general education outcomes and has data available for more than one class of graduates during the last three years. | 10 |
| (3) The institution has assessed the performance of its graduates on a measure of general education outcomes and for at least one assessment during the last three years can demonstrate that its graduates performed equivalent to graduates from similar institutions. | 15 |
| (4) The institution has assessed the performance of its graduates on a measure of general education outcomes and for at least one assessment during the past three years can demonstrate that its graduates performed above graduates of similar institutions. | 20 |

* Graduates for its major degree—associate for two-year institutions and bachelor's for senior institutions.

** General Education outcomes are generally defined here as performance on major intellectual skills and knowledges expected of graduates with a particular degree—communication, problem solving, reasoning, familiarity with major modes of thought, etc. The measure of outcome must be a nationally prepared assessment instrument having norms beyond the institution. Examples would include the ACT COMP battery, the ETS GRE Aptitude tests, the ETS Undergraduate Assessment Program, or the ETS Test of Academic Competencies and General Education.

NOTE. The report must specify the instrument, the time/date of administration, the population or sample assessment, data results and analysis, comparison with some comparative or absolute standard.

TABLE *(continued)*

**Performance Variable IV—Evaluation of Institutional Programs and Services by Enrolled Students, Recent Alumni, and Community Members/Employers**

| Standards of Performance | Suggested Point Allocation |
|---|---|
| For any year in the past three years, in institution has conducted a survey of referent group evaluation of institutional programs and services.* The three referent groups are currently enrolled students, recent alumni, and community members/employers. | |
| (1) For one of these referent groups with limited program or service application. For example, a survey of graduates for one or two academic fields. | 5 |
| (2) For two or more of these referent groups with limited program or service application. | 10 |
| (3) For one of these referent groups with application to entire institution. | 15 |
| (4) For two or more of these referent groups with application to entire institution. | 20 |

*The survey instrument employed may be a nationally constructed instrument such as the Student Reaction to College published by ETS, the NCHEMS Program Completer Questionnaire, or a locally constructed instrument. The instrument must yield quantifiable responses which reflect satisfaction or evaluation indices. The report must describe the instrument, time/date of use, the population surveyed, the response rate, summary and analysis of data, and policy program improvement actions taken as a result.

TABLE *(continued)*

**Performance Variable V—Peer Evaluation of Academic Programs**

| Standards of Performance | Suggested Point Allocation |
|---|---|
| (1) The institution has conducted a formal evaluation of at least two major program fields during the last three years—an evaluation utilizing a peer review team of scholars from other institutions outside the state and/or practicing professionals within a field.* The institution can further indicate what changes in policy or practice were made as result of evaluations. | 5 |
| (2) The institution can demonstrate that it has conducted at least five evaluations within the last five years as part of *an ongoing program of peer review* and can further indicate changes in policy or practice made as a result of these evaluations. | 6-15 |
| (3) The institution can demonstrate that one or more of its academic programs enjoy a favorable peer reputation beyond the state. Evidence of such peer reputation can take the form of national peer reviews similar to the Cartter study or the Roose Anderson study of graduate programs or institutionally conducted studies. | 15-20 |

* A peer review team must consist of two or more persons who have visited the institution.

NOTE: The supporting data for this variable must describe the program field; the names, position, and brief vita outline of peer team members; the dates of visit/evaluation; criteria utilized; the major findings; and follow-up actions taken.

TABLE *(continued)*

**Performance Variable VI—Performance of Students or Evaluation of Instructional Programs Not Included in Other Variables**

| Standards of Performance | Suggested Point Allocation |
|---|---|
| (1) Institution has assessed performance* of a representative sampling of students** or otherwise assessed the effectiveness of an instructional project within the last three years. | 5 |
| (2) Institution has assessed performance of a representative sampling of students for each of the past three years or has assessed the effectiveness of one or more instructional projects within each of the last three years. | 10 |
| (3) Institution has assessed performance of a representative sampling of students within the past three years and can demonstrate that the performance of its students ranks with or above the performance of students from similar institutions or has assessed the effectiveness of one or more instructional projects and can demonstrate that the effectiveness of said project(s) exceeds that of similar projects at other institutions. | 12-20 |

*The measure of performance must be an assessment instrument/procedure constructed external to the institution—with normative standards available for state, regional, or national samples.

**The number of students assessed must be sufficiently representative to permit statistically sound inferences to all students in that field.

APPENDIX B

# Instructional Evaluation Variables and Standards 1981-82

Revised Draft
April 1980

| Variable | Weight |
|---|---|
| I. Program Accreditation | 20 |
| II. General Education Outcomes | 20 |
| III. Program Performance Outcomes | 20 |
| IV. Program/Services Satisfaction Indices | 20 |
| V. Evaluation Planning/Action for Renewal and Improvement | 20 |

## I

## Instructional Evaluation Variable—Revised

## Program Accreditation

| Performance Standard | Points Awarded |
|---|---|
| An institution will be awarded points on this variable based on the percentage of eligible programs accredited. For example, an institution having 30% of its eligible programs accredited would be awarded 6 points. An institution having 50% accredited would be awarded 10 points. | 0-20 |

### Definitions, Commentary, Procedures

(1) A "program" is defined as a sequence of courses and/or other educational experiences leading to a degree major as carried in the THEC program inventory.

(2) A program is defined as "eligible" if there is a COPA approved agency or organization which accredits programs for that field and degree level.

(3) Program fields covered by an umbrella accreditation will be counted as "one" unit. For example, if an institution offers five bachelor's degree majors in business, and the business school or college is AACSB accredited at the undergraduate level, these five programs will be counted as one program for purposes of this variable.

(4) Programs automatically excluded from the list of eligible programs are programs (a) that have been approved by THEC for less than five years, (b) that are being terminated or phased out—based on official action of the appropriate governing board and formal notification to the Commission of such action, and (c) that have been officially identified by the respective governing boards as "inactive" and formal notification furnished to the Commission for its inventory records.

(5) A program eligible for accreditation by more than one agency will be counted only once in the "eligible" list.

(6) THEC staff will prepare a list of eligible programs from program inventory records. This list will serve as the official list of eligible programs unless institutions request and THEC formally approves the exemption of a program. Such THEC action of exception must take place at least one meeting prior to the meeting each year in which the appropriations recommendations are adopted by the THEC.

## II

## Instructional Evaluation Variable

## General Education Outcomes

| Performance Standards | Points Awarded |
|---|---|
| (1) The institution has assessed the performance of a representative sampling of graduates for its major degree—associate or bachelor's—on a measure of general education outcomes at least once during the past four years. | 5 |
| (2) The institution has, during the last four years, assessed the general education performance of a representative sampling of a majority of its graduates by major field or college, and has begun a program of inter-field or inter-collegiate analyses of the data. | 10 |
| (3) The institution has an ongoing program to assess the performance of its graduates on a measure of general education outcomes and has available data, preferably on the same measure, for representative samples of two or more classes of graduates during the previous four years. | 15 |
| (4) The institution meets the requirement of standard (3) and can further demonstrate for the most recent or one of the two most recent assessments that the development of its graduates—that is, the change in performance from freshman to graduation—is equivalent to or greater than the development of students from at least one institution whose freshmen performance is at a comparable level. | 20 |

### Definitions, Commentary, Procedures

(1) "General Education Outcomes" are generally defined as that knowledge and those skills expected of graduates earning the major degree of an institution. These may include communication, problem solving ability, reasoning skill, analytic and synthesis skills, familiarity with major modes of inquiry, etc. The specific definition of these outcomes is expected to reflect the mission, philosophy, and special character of each institution.

(2) The "measure of outcome" must be an assessment instrument having norms beyond the institution. Examples would include the ACT COMP battery, the ETS GRE Aptitude tests, and ETS Undergraduate Assessment Program, the Adult Proficiency Level Examination, elements of the National Assessment of Education Progress. This list of examples is not intended to be exhaustive.

(3) A "representative sampling" is defined as a sample of graduates chosen so that the sample statistically represents the population of graduates. The population of graduates is presumed to include all those receiving the institution's major degree for a given year.

(4) An "ongoing program" of general education assessment is defined as a program described in formal institution policy and published in appropriate academic policy documents.

(5) Information supplied in support of performance on this variable and its standards should include:

(a) A brief description of the instrument employed and the agency or company publishing the instrument. The general education outcomes assessed by the instrument should be concisely described.

(b) The dates of administration.

(c) A description of the population or sample assessed—including size of the sample and other evidence of how the sample was chosen to represent the population.

(d) A concise presentation and analysis of results for each administration.

(e) An analysis of those institutions and/or referent student populations judged to be comparable for those institutions attempting to qualify on standard (3). The analysis should include the data basis for concluding that the institution's graduates developed at a rate equivalent to or greater than students from comparable institutions.

(f) **A description and analysis of instructional policy or practice changes made as a result of institutional review of the data obtained from the general education assessment after two years' data have been collected.**

## III

## Instructional Evaluation Variable

## Program Performance Outcomes

| Performance Standards | Points Awarded |
|---|---|
| (1) Within the past five years, the institution has assessed the performance of a representative sampling of graduates in program fields leading to its major degree—associate or bachelor's. Points shall be awarded on the percentage of eligible fields. For example, an institution having assessed 50% of its eligible fields would be awarded 5 points. | 0-10 |
| (2) The institution can meet the requirement of standard (1) and has an ongoing program to assess the performance of its graduates in a majority of its major program fields. For each program field reported in standard (1), the institution has data available, preferably on the same measure, for a representative sampling of two or more classes of graduates within the past five years. | 15 |
| (3) The institution meets the requirements of standard (2) and can further demonstrate that the performance of its graduates in the majority of those program fields assessed is above the performance of graduates from the same field in comparable institutions. | 20 |

### Definitions, Commentary, Procedures

(1) A "program field" is defined as a sequence of courses and/or other educational experiences leading to a degree major as carried in the THEC program inventory.

(2) "Performance" is defined as the scores of students on an assessment instrument/procedure constructed external to the institution and having normative standards for state, regional, or national referent groups. Examples would include the GRE field tests, state or national licensing examinations, professional field tests such as National Teacher Examination, Engineer in Training Examination, etc.

(3) An "eligible" field is one in which there is an assessment instrument available and which field is carried in the THEC Academic Inventory. For purposes of this variable, "eligible" fields may be clustered for related majors. For example, program majors in Accounting, Management, Marketing, etc., may be counted as one "eligible" program if they are assessed by a common instrument.

(4) An "ongoing program" to assess the performance of graduates in major program fields is defined as a program described in formal institutional policy and published in appropriate academic policy documents.

(5) A "representative sampling" of graduates is one chosen so that it statistically represents the population of graduates in a particular program field/major.

(6) A "comparable" institution is one whose entering freshmen performance aptitudes are similar.

(7) The following material should be supplied for each eligible program major.

(a) The program or professional field assessed.

(b) A brief description of the instrument employed, including the firm or agency publishing the instrument.

(c) The date(s) of administration during the past five years (from September 1, 1975 through summer 1980).

(d) A brief presentation and analysis of results.

(e) If comparisons with comparable institutions are used to qualify on standard (3), the analysis should include the data basis for concluding that performance is equivalent to or greater than comparable institutions.

(f) A description and analysis of instructional policy or practice changes made as a result of institutional review and evaluation of the data obtained from these assessments after two years' data have been collected.

## IV

## Instructional Evaluation Variable

## Programs/Services Satisfaction

## Indices

| Performance Standards | Points Awarded |
|---|---|
| During the past four years (1976-77, 77-78, 78-79, 79-80), an institution has conducted evaluative surveys of instructional programs/services for a representative sampling of *currently enrolled students, recent alumni,* or *community members/employers.* | |
| (1) Institution has surveyed *at least one* of the three referent groups with application to the majority of its program fields or to the entire institution. | 10 |
| (2) Institution has surveyed two or more of the referent groups with application to the majority of its program fields or to the entire institution | 20 |
| or | |
| Institution has evaluation/satisfaction data available from *two or more surveys of the same referent group* utilizing the same survey instrument—with application to majority of program fields or entire institution. That is, the institution has an ongoing assessment of its programs/services with a recurring or periodic survey of the same referent group and can present comparable evaluation data from at least two such surveys within the past four years. | |

### Definitions, Commentary, Procedures

(1) An "evaluative survey" is defined as one yielding quantifiable indices reflecting satisfaction or evaluation of instructional programs or services. The survey instrument may be a nationally or locally constructed instrument. Examples would be the NCHEMS Program Completer questionnaires, the ETS *Student Reaction to College,* the UCLA *College Student Experiences Questionnaire,* ACT *Evaluation/Survey Services.*

(2) A "program field" means a sequence of courses and/or educational experiences leading to a degree major as carried in the THEC program inventory. For purposes of this variable, a program field may also designate a cluster of related majors—such as programs in business, allied health, education, engineering, humanities, etc.

(3) "Application to majority of program fields" is intended to describe a survey which applies to more than half of individual majors or academic units. If an institution offers 60 majors in five academic units, a "majority of program fields would mean surveys applying to 31 or more of the majors, or surveys conducted by at least three of the five academic units.

(4) A representative sampling means a sample so chosen that it statistically represents the population. The response rate is expected to approximate that of other similar surveys, to approach 50% or greater.

(5) Data to be provided in support of this variable should include the following:

(a) A brief description of the survey instrument employed. The description should indicate the company or firm publishing the instrument (if a local instrument was not employed) and concisely outline which program, service, or policy factors were evaluated. A copy of the survey may be included—and must be included if an institutionally constructed instrument was employed.

(b) The date(s) of administration during 1976-77, 1977-78, 1978-79, 1979-80 years.

(c) A description of population or sample surveyed and *response rate*, and (if appropriate) methods used to check non-response bias.

(d) A brief presentation and analysis of results.

(e) **A description of specific policy or program improvement actions taken as a result of the survey.**

## V

## Instructional Evaluation Variable

## Evaluation Planning for Renewal and Improvement

| Performance Variable | Points Awarded |
|---|---|
| This variable encourages the development of a comprehensive institutional evaluation plan centering instruction on improvement. It also recognizes that institutions will have designed and implemented some evaluation activities which make major contributions to instructional renewal and improvement but these activities may not be adequately reflected in any previous four variables. Institutions may submit activities that . . . | 0-20 |

(1) Have been designed and implemented within the past four years.

(2) Have yielded at least one set of evaluation results. Activities planned but not yielding results are not eligible.

(3) Have direct impact on the effectiveness of educational programs, services (advising, etc.), or faculty/administrative performance.

Evaluation accomplishments submitted on this variable will be reviewed against these criteria:

(1) The extent to which the activity is part of a comprehensive institutional evaluation plan.

(2) The extent to which the activity complements rather than duplicates evaluation activities of the first four variables.

(3) The extent to which the activity involves the endorsement and participation of institutional faculty.

(4) The extent to which the activity accents the mission of the institution.

(5) The extent to which the activity supports goals-objectives of its governing board and those of the THEC 1979 Master Plan.

(6) The extent to which the activity involves the application of or reference to judgments and/or performance standards external to the institution.

(7) The extent to which the activity reflects the application of imaginative or new ideas/approaches to evaluation.

(8) The extent to which the activity directly affects instructional policy or practice.

(9) The extent to which the activity appears to have recognition beyond the institution.

(10) The extent to which results and analysis are reported in specific and clear style.

APPENDIX C

# Instructional Evaluation Variables
# November 21, 1983
# Tennessee Higher Education Commission

---

## I

## Instructional Evaluation Variable

## Program Accreditation

### Purpose

This variable is designed so as to reward institutions that design and offer academic programs, for which accreditation services are provided, that meet or exceed the standard of responsible accreditation agencies.

### Performance Standard and Point Allocation

An institution may be awarded up to 25 points on this variable. The number of points awarded to the institution will be a percentage of this maximum amount calculated as the percentage of eligible programs accredited.

### Definitions and Procedures

(1) A "program" is defined as a sequence of educational experiences leading to a degree major as listed in the THEC program inventory.

(2) A program is defined as "eligible" if there is a COPA member agency or organization which accredits programs for that field and degree level (unless exempted under (6) below). Additional accrediting agencies may be proposed by governing boards. Upon THEC staff approval, all programs accreditable by such agencies will be included as eligible statewide.



139

(3) Program fields covered by an umbrella accreditation will not be counted as "one" unit, but each degree major as "one." For examp e, if an institution offers five bachelor's degree majors in business, and the business school or college is AACSB accredited at the undergraduate level, these five programs will be counted as five programs for the purposes of this variable.

(4) Programs automatically excluded from the list of eligible programs are programs (a) that have been approved by the THEC for less than five years, unless the program is accredited by a COPA agency, (b) that are being terminated or phased out—based on appropriate official action, and (c) that have been identified as inactive by the appropriate board and the THEC.

(5) A program eligible for accreditation by more than one agency will be counted only once on the eligible list.

(6) Where program accreditation efforts are shown to be unjustified on a statewide basis in relation to an accumulation of factors such as economic feasibility, critical mass of enrollees, low benefits to students, more important qualitative priorities, etc., institutions may request respective governing board to seek program exception. Any exception approved by the THEC staff must apply to all similar program areas in the state.

(7) Proposals from governing boards for statewide changes in eligibility of programs or appropriateness of accrediting agencies as outlined in (2) and (6) above must be submitted to the THEC staff before January 1 each year to facilitate any necessary revision of the eligible program or acceptable accrediting agency lists for the next budget cycle. The official list of eligible programs or appropriate agencies shall be maintained by the THEC staff based on inventory records and approved exceptions as noted above.

## II

## Instructional Evaluation Variable

## Program Field Evaluation

### Purpose

This variable consists of two standards. An institution may earn a maximum of 10 points under the first standard (IIA) and a maximum ot 20 additional points under the second standard (IIB). The first standard is intended to encourage an institution to evaluate the quality of each of its academic programs at least once within a five-year period. The second standard is designed to reward those institutions which can demonstrate on the basis of test results that the quality of their programs is increasing or has attained an above average level of quality. Together, these standards provide a means of evaluating the quality of the specialized academic offerings of institutions.

### IIA Program Field Evaluation

### IIA

### Externally Validated Tests, Locally Developed Test, or External Peer Review

### Performance Standard and Point Allocation

Under Standard IIA, an institution may be awarded up to 10 points. The number of points awarded to the institutions will be a percentage of this maximum amount calculated as the percentage of program fields which have met the requirements outlined below within the past five academic years.*

The institution has assessed the performance of a representative sampling of graduates of the program field by means of an externally validated instrument approved by the THEC staff. This instrument shall be applied to and appropriate for the program level which has produced the largest number of graduates in recent years at that institution.

OR

The institution has assessed the performance of a representative sampling of program field graduates by process of the administration of a locally developed program test. This instrument shall by applied to and appropriate for the program level which has produced the largest number of graduates in recent years at that institution.

OR

The institution has evaluated the quality of individual programs via external peer review (this alternative is not available for accreditable programs). This evaluation shall cover all levels of the program field offered by the institution.

* For the first four years' administration of this variable, the 10 points maximum will be awarded to an institution according to the following schedule.

  A. In the first year—at least 20% of the program fields have met one of the listed requirements within the first year.

  B. In the second year—at least 40% of the program fields have met one of the listed requirements within the first or second year.

  C. In the third year—at least 60% of the program fields, have met one of the listed requirements within the first, second, or third year.

  D. In the fourth year—at least 80% of the program fields have met one of the listed requirements within the first, second, third, or fourth year.

## IIB Program Field Evaluation

### Improved Programs or Programs of Exceptional Quality

### Performance Standard and Point Allocation

To be eligible for points under this standard, an institution must demonstrate that it has assessed the performance of a representative sample of graduates of its program fields via externally validated tests or locally developed tests. Up to 20 points may be awarded under Standard IIB. The number of points awarded to the institution will be a percentage of this maximum calculated on the basis of the percentage of programs that have met the requirements outlined below within the past five academic years. (See the table below for further details.)

The institution can demonstrate on the basis of an externally validated test appropriate to that field that the performance of program graduates exceeds the norm.

OR

The institution can demonstrate on the basis of externally validated test appropriate to that field that the performance level of program graduates exceeded the level of performance by program graduates on the most recent administration of that test.

OR

The institution has assessed the performance of a representative sampling of program graduates through administration of a locally developed test and can demonstrate program graduate scores which exceed the scores from the most recent previous administration of that test.**

**Awards Under Standard IIB**

| Percentage of Program Fields Meeting Requirements | Points Awarded |
|---|---|
| 75% — 100% | 20 |
| 72.5% — 74.9% | 19 |
| 70.0% — 72.4% | 18 |
| 67.5% — 69.9% | 17 |
| 65.0% — 67.4% | 16 |
| 62.5% — 64.9% | 15 |
| 60.0% — 62.4% | 14 |
| 57.5% — 59.9% | 13 |
| 55.0% — 57.4% | 12 |
| 52.5% — 54.9% | 11 |
| 50% —52.4% | 10 |
| 47.5% — 49.9% | 9 |
| 45.0% — 47.4% | 8 |
| 42.5% — 44.9% | 7 |
| 40.0% — 42.4% | 6 |
| 37.5% — 39.9% | 5 |
| 35.0% — 37.4% | 4 |
| 32.5% — 34.9% | 3 |
| 30.0% — 32.4% | 1 |
| 27.5% — 29.9% | 1 |
| 0 — 27.5% | 0 |

** In order to compensate institutions for the initial costs of developing local tests, institutions will be rewarded for the first administration of such tests during the first five years (to July 1, 1988). A locally developed test administration for the the first time in this period will be scored as if the institution's test scores had exceeded a previous score on the same test.

## Definitions and Procedures

(1) In general, a "program field" is defined as all levels of programming bearing the same name as an academic major. A group of closely related programs with dissimilar names may also be considered a single program field. General technology and general transfer programs leading to an associate degree are exempt from this variable as are pre-professional programs which do not result in a degree under

that name. All individualized programs offered by an institution will count as one program field for purposes of this variable. Programs automatically excluded from consideration under this variable are programs (a) that have been approved by the THEC for less than five years, (b) that are being terminated or phased out—based on appropriate official action, (c) that have been identified as "inactive" by the appropriate board and the THEC, and (d) that are offered at a level below the baccalaureate at baccalaureate degree granting institutions (nursing programs excepted). Program fields which are accredited under Variable I and which are performance oriented shall not be included in this variable. A list of program fields for each institution shall be maintained by the THEC staff.

(2) An institution choosing to conduct external peer reviews must submit a plan for external review through its governing board staff to the THEC staff for approval prior to the review. The plan for external peer review must include names and vita of at least two peers proposed to conduct the review, a schedule of planned activities to be included in the review, include efforts to measure the improvement of educational outcomes to the maximum extent possible, and provide for a written critical report summarizing the findings which will be forwarded to the THEC as part of the budget request process.

(3) A "locally developed program test" must be constructed in cooperation with at least one similar institution with a similar degree major program or in consultation with a team of at least two external consultants, at least one of which must be an expert in the content of that program field.

(4) An institution choosing to use locally developed program tests must submit a plan for test construction through its governing boards of staff for THEC staff approval prior to construction. The plan for test construction must include a schedule of activities, sampling procedure, credentials of cooperating institution staff or credentials of external consultants, and a proposed schedule for submission for THEC staff approval prior to use. Results and analysis of locally developed program tests must be submitted as part of the budget request process.

(5) The master list of appropriate externally validated tests available for programs will be determined and maintained by the THEC staff.

(6) In choosing among externally validated tests, locally developed tests or peer reviews, an institution should consult its governing board staff.

(7) In reporting test results under Standard IIB utilizing program field tests for which more than one datum descriptive of average performance for both the sample of graduates and the norm group is available, a mean score will be accepted in preference to a percentile datum and a

percentile datum will be accepted in preference to a pass/fail rate. If more than one norm group is available for comparison, national norms are preferred to regional norms which are preferred to state norms.

(8) The awarding of points under both IIA and IIB is on the basis of "official test" scores. Once a test is given and points awarded under IIA, the score reported becomes the "official test" score for five years or until an institution notifies the THEC staff that it intends to "retest" that field within the five-year period. Intent to retest must be declared in advance of such testing and the results must be reported to the THEC. The score of the retest becomes the "official test" score and is utilized for all point calculations under variable II.

(9) Reference to the "most recent previous administration" of a test refers to the most recent administration of an "official test" (defined above). An exception to this is made for the first test of a program field qualifying as an "official test."

## III

## Instructional Evaluation Variable

## Institution-wide Education Outcomes

### Purpose

This variable consists of two alternative standards. The particular standard to be applied is dependent on the class of institution. This variable provides a means of evaluating the general (non-program-specific) quality of the educational program at each institution.

### IIIA General Education Outcomes

This standard shall apply to all four-year institutions and may apply to community colleges as described below.

### Performance Standards and Point Allocation

(1) The institution will be awarded 5 points if, within the past five academic years, the institution has assessed the performance of a representative sample of graduates for its major academic degree utilizing the ACT-COMP Objective or Composite measure.

(2) The institution will be awarded an additional 20 points if through annual assessment utilizing the ACT-COMP measure, the institution can demonstrate that the performance of its graduates regarding value added is above average when compared with the performance of graduates of comparable institutions. (See definition #6 for procedure.)

OR

The institution will be awarded an additional 20 points if through annual assessment utilizing the ACT-COMP measure, the institution can demonstrate an improvement in value-added from the most recent institutional measure of value-added. (See definition #6 for procedure.)

### IIIB Placement of Graduates

This standard shall apply to all technical institutes. Community colleges must first make a determination as to which of their programs can be assessed by a measure of general education outcomes. For these programs, the standards of IIIA shall apply. For the remaining programs at these institutions, IIIB shall apply. The total number of points awarded shall be prorated between the two standards according to this division.

### Performance Standards and Point Al ɔcation

(1) For programs being evaluated within this standard, the institution will be awarded 5 points if the institution each year has conducted a follow-up survey of all graduates to ascertain their employment status in the cluster of occupations for which they were trained.

(2) The institution will be awarded an additional 20 points if, through analysis of the surveys conducted in IIIB(1), the employment rate for graduates in the cluster of occupations for which they were trained exceeds 70 percent.

OR

The institution will be awarded an additional 20 points if the employment rate of graduates in fields for which they were trained exceeds the employment rate in the most recent similar survey of employment rate of graduates in fields for which they are trained.

### Definitions and Procedures

(1) Follow-up surveys must be designed to establish the employment status of all program completers during a period not earlier than 30 days following program completion and not later than 90 days following program completion. The single exception to this shall be the survey of students completing in the spring quarter in time for a May or June convocation. These students must be surveyed not earlier than September 1 and not later than October 31 following their program completion.

(2) All completers surveyed within a fiscal year will form the basis of calculation of employment rate. The placement percentage is calculated as the ratio of the total number of students placed in fields for which they were trained to the total number of program completers less those in military service or pursuing further education.

(3) A list of "clusters of occupations" appropriate to each program subject to evaluation under standard IIIB shall be maintained by the THEC staff.

(4) A representative sample is a sample of entering students or graduates chosen so that the sample statistically represents the population of entering students or graduates for a given year.

(5) Value added shall be measured by a comparison of the general education mean score as measured by the ACT or COMP for entering freshman to the mean COMP score for a graduating class. Any one of the following procedures may be used:

(a) Longitudinal Study using the COMP Composite Examination
(b) Longitudinal Study using the COMP Objective Test
(c) Cross-sectional Study using the COMP Composite Examination
(d) Cross-sectional Study using the COMP Objective Test
(e) Exit-level assessment only, estimating the entry level COMP score based on a concordance table with the ACT composite score.

(6) Above average performance in value-added must be demonstrated by an institution having a value-added mean score which exceeds the value-added mean score for a similar set of institutions measuring value-added with a comparable procedure. A similar set of institutions shall number no less than six and shall include, to the extent possible, institutions with similar purposes, similar enrollments, similar support systems, and similar testing or surveying techniques. A similar set of institutions cannot be exclusively or predominantly composed of in-state public institutions. The THEC staff shall determine which set of institutions are to be considered similar following consultation with institutional and governing board staffs and with personnel from the American College Testing Program.

(7) The sampling procedure for activities in this variable must be submitted prior to use for THEC staff review.

(8) Institutions must submit a written report including scores, survey results and analyses as part of the budget request process.

(9) Calculations of value added shall be rounded to the nearest whole number and point allocations made on that basis.

## IV

## Instructional Evaluation Variable

## Instructional Improvement Based on Referent Group Survey

### Purpose

This variable consists of two standards. A maximum of 5 points can be earned under each standard for a total of 10 points under this variable. This variable is designed to encourage institutions to seek evaluation of its overall academic program quality by consumers of the educational product.

### IVA Single Survey

### Performance Standards and Point Allocation

The institution will be awarded 5 points upon demonstration that the institution has surveyed, with an evaluative instrument, a representative sample of at least one of three referent groups (enrolled students, formerly enrolled students, or community members/employers) with application to the majority of its program fields or to the entire institution. To be awarded points for this standard the institution must submit a brief presentation of the analysis of the survey results and provide a description of specific substantial, instructional improvement actions taken as a result of the survey and analysis when such improvement actions are indicated.

### IVB Two Surveys

The institution will be awarded an additional 5 points if the institution has surveyed, with an evaluative instrument, two or more of the referent groups with application to the majority of its program fields or to the entire institution. To be awarded points for this standard, the institution must submit a brief presentation of the evaluative survey results and provide a description of specific, substantial, instructional improvement actions taken as a result of the surveys and analyses when such improvement actions are indicated.

OR

The institution will be awarded an additional 5 points if the institution can demonstrate that an evaluative survey has been administered more than once to the same referent group and can demonstrate for this referent group's most recent evaluation, the institution has received improved scores from the previous survey taken as a whole.

## Definitions and Procedures

(1) An "evaluative survey" is defined as one yielding quantifiable indices reflecting satisfaction with or evaluation of instructional programs. The survey instrument may be a nationally or locally constructed instrument. A list of acceptable instruments for this variable will be maintained by the THEC staff. Prior approval by the THEC staff for the use of instruments not on this list is required.

(2) A representative sample means a sample chosen so that it statistically represents the population.

(3) Instructional improvement actions must relate directly to improvement of classroom instruction or indirectly in terms of academic support activities such as library services, academic counseling services, etc. (items such as food service, parking, or other student conveniences, etc., are excluded).

(4) As part of the budget request, an institution must submit a copy of the survey instrument, date(s) of administration, description of sampling procedure, and analysis sufficient for any points claimed.

(5) To qualify as a survey, it is not necessary that a single instrument be used. Multiple instruments employed within the same fiscal year constitute a survey when, taken as a group, they are applied to a majority of the institutional program fields or to the entire institution.

(6) To be awarded points under this variable, the survey or surveys must be conducted during the fiscal year immediately preceding the fall appropriations request cycle.

# V

## Instructional Evaluation Variable

## Planning for Instructional Program Improvement

### Purpose

This variable consists of two standards. A maximum of 5 points can be awarded under each standard for a total of 10 points under this variable. This variable is designed to encourage institutional self-evaluation of its academic program quality.

### VA Plan Development

### Performance Standards and Point Allocation

The institution will be awarded 5 points under this standard, provided it submits an acceptable annual plan for instructional improvement in the forthcoming year to the THEC staff at the time of submission of appropriations requests for that forthcoming year. An acceptable plan must exhibit these features:

a. Specific goals and benchmarks or measurable objectives can be reached in the planning period are set forth.

b. Activities scheduled as part of the plan must provide for an evaluation component.

c. All activities which form the basis of claims for points under the other four institutional evaluation variables should be included in the plan, but the plan should address additional means of instructional improvement.

d. Faculty must be involved in the development, execution, and evaluation of the plan.

e. The plan must be focused upon improvement in instruction, either directly in terms of improved classroom performance as illustrated by outcomes measures or indirectly in terms of improvement to academic support activities such as library services, academic counseling services, etc.

f. The plan should be consistent with longer term plans of the institution, its governing board, and the THEC.

## VB Plan Evaluation

### Performance Standards and Point Allocation

The institution will be awarded an additional 5 points under this standard, provided it submits at the time of submission of appropriations requests an evaluation of the plan for instructional improvement covering the previous year. This evaluation must report the degree to which the plan was executed and the results obtained in terms of reaching goals and benchmarks or measurable objectives and completion of activities. Only those institutions which can demonstrate that at least half of the objectives and benchmarks have been reached or activities favorably evaluated will be awarded points under this standard.

APPENDIX D

# Instructional Evaluation Information for College D

*Prepared for*

College D

*by*

The Learning Research Center
1819 Andy Holt Avenue
Telephone 974-2459

153

## Sources of Instructional Evaluation Information for College D

Note: (+) or (−) Based on University Average;
(0) = Lowest rating of all Colleges

### I. Student Achievement

General Education
- Mean Entering ACT Score (+)
- Mean Total COMP Score (+)
- Estimated Gain on COMP (−)

Achievement in Major
- National Professional Exam

### II. Other Sources

- Dean's Follow-Up Survey of Seniors (Winter 1984)
- Student Evaluations of Instruction

### III. Student Ratings

Quality of Program Services in the Major—Table 12

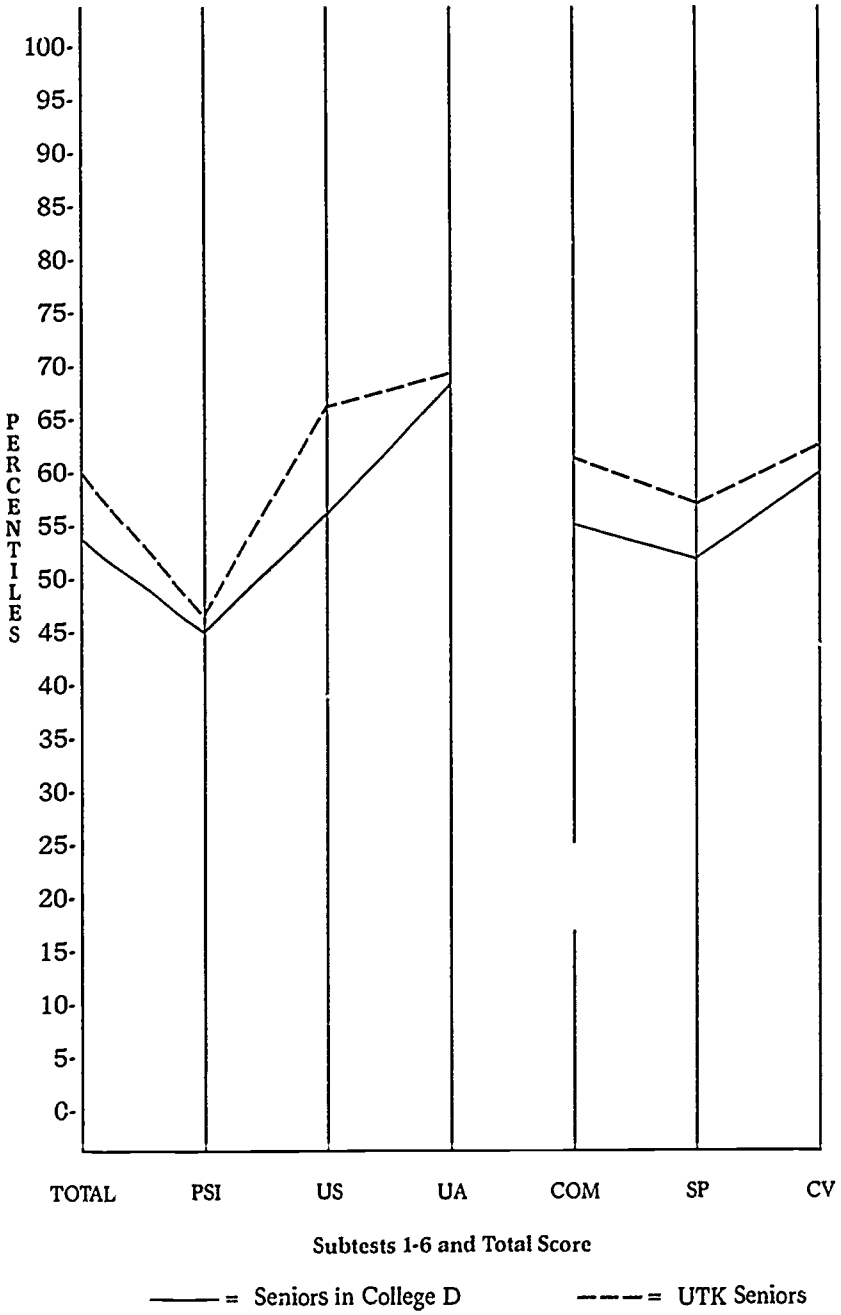
| | |
|---|---|
| Availability of Advisor | (0) |
| Willingness of Advisor to Help | (−) |
| Quality of Printed Program Information | (−) |
| Helpfulness of the Office Staff | (0) |
| Quality of Special Events | (−) |
| Adequacy of Preparation by Lower Division Courses for Upper Division Courses | (0) |
| Quality of Courses: Providing General Education | (−) |
| Quality of Courses: Preparing for Employment | (0) |
| Availability of Required Courses for the Major | (−) |
| Availability of Desired Courses for the Major | (−) |
| Organization of the Curriculum | (−) |
| Fairness of Grading | (−) |
| Quality of Instruction in Lower Division Courses in the Major | (0) |
| Quality of Instruction in Upper Division Courses in the Major | (−) |
| Opportunities for Interaction with Faculty in the Major | (0) |
| Practicum/Intern in the Major | (+) |
| Library Collection Related to the Major | (−) |

Quality of Classroom Experience—Table 15

| | |
|---|---|
| Comprehensiveness of Course Content | (−) |
| Relevance of Content for Student Needs | (0) |
| Extent to which Content Is Current | (−) |
| Instructor's Class Presentations | (0) |
| Instructor's Class Preparation | (−) |
| Instructor's Enthusiasm for Teaching This Class | (−) |
| Instructor's Help with Problems | (−) |
| Fairness: Testing | (0) |
| Fairness: Grading | (−) |
| Clarity of Course Objectives | (−) |
| Conduciveness of Climate: Learning | (0) |
| Relevance of Lecture Information to Course Objectives | (0) |
| Quality of Classroom Discussion | (−) |
| Accuracy of Catalog Description: Course | (0) |
| Instructor's Knowledge of Subject Matter | (−) |
| Instructor's Availability for Consultation | (0) |
| Overall Quality of Instructor | (−) |
| Overall Quality of Course | (−) |

| | |
|---|---|
| Satisfaction with Social Experience at UTK | (+) |
| Satisfaction with Academic Experience at UTK | (+) |

## COMP Results Profiled on Senior Reference Group Norms Based on Percentiles for 8510 Seniors at 70 Institutions



## University of Tennessee—Knoxville COMP Objective Test Means (Equated to Form III) for Senior Samples Tested in 1980-1983

| | May 1980 N = 165 Seniors | April 1981 N = 680 Seniors | April 1982 N = 644 Seniors | May 1983 N = 700 Seniors | Maximum |
|---|---|---|---|---|---|
| Total | 192.3 | 187.9 | 189.3 | 188.9 | 240 |
| FSI | 66.1 | 65.1 | 64.1 | 63.5 | 80 |
| UC | 66.1 | 64.0 | 64.3 | 64.5 | 80 |
| UA | 60.1 | 59.1 | 61.2 | 61.3 | 80 |
| COM | 55.8 | 54.7 | 54.3 | 54.0 | 72 |
| SP | 78.3 | 76.5 | 76.7 | 76.6 | 96 |
| CV | 58.2 | 56.8 | 58.2 | 58.2 | 72 |
| ACT/E | 23.6 | — | 21.5 | 22.3 | |
| (N) | (119) | — | (467) | (382) | |

## COMP Scales

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

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

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

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

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

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

## University of Tennessee, Knoxville
## COMP Estimated Mean Gains by College
## Based on Seniors with ACT Scores

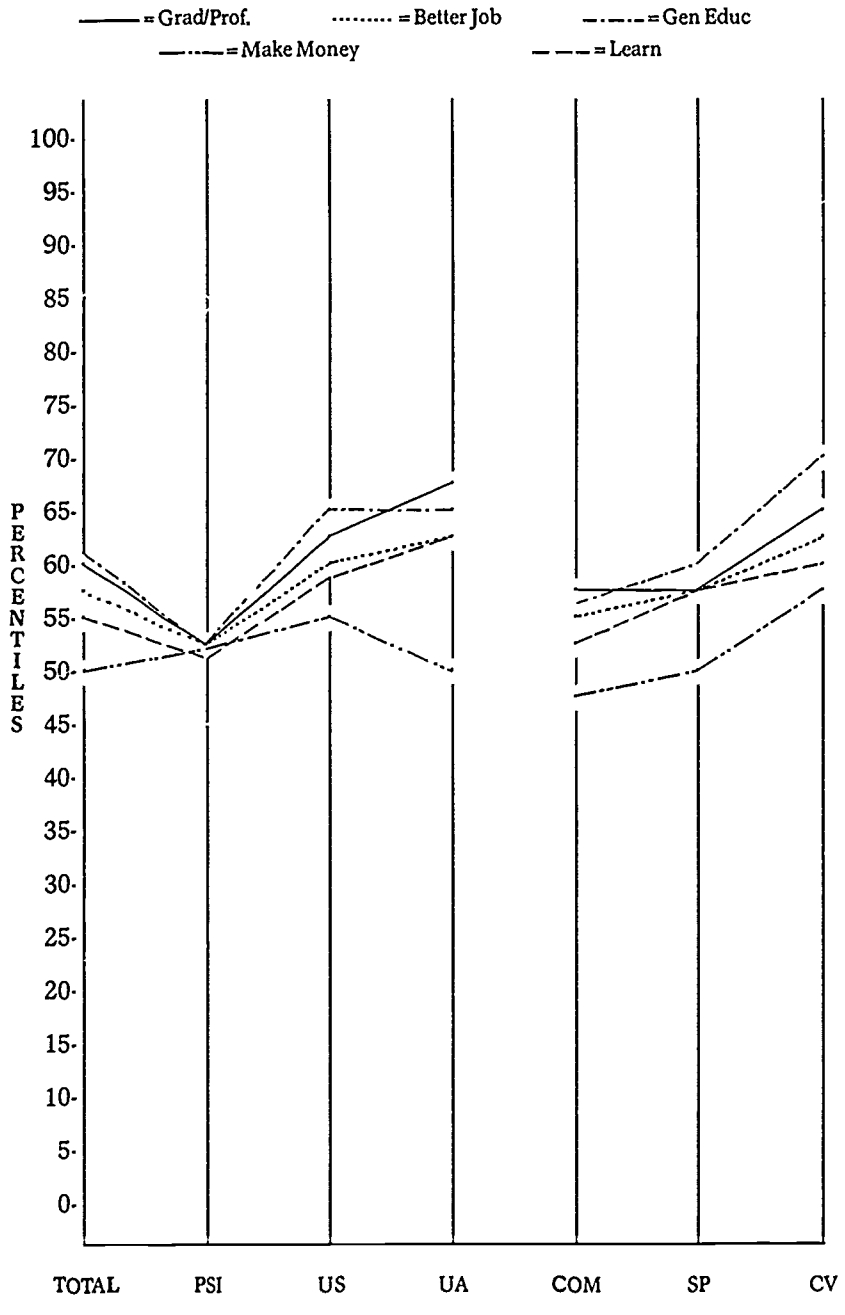
| College | 1982 | | | 1983 | | |
|---|---|---|---|---|---|---|
| | N/ACT Score | Mean ACT Composite | Estimated Mean Gain on COMP | N/ACT Score | Mean ACT Composite | Estimated Mean Gain on COMP |
| A | 32 | 22.1 | 7.2 | 29 | 22.1 | 9.3 |
| B | 8 | 26.0 | 6.4 | 17 | 23.8 | 4.1 |
| C | 142 | 21.3 | 9.0 | 101 | 21.7 | 6.1 |
| E | 34 | 19.9 | 9.0 | 35 | 20.9 | 10.1 |
| F | 42 | 18.2 | 13.2 | 25 | 17.2 | 11.1 |
| D | 56 | 25.6 | 2.5 | 78 | 24.3 | 1.7 |
| G | 63 | 19.8 | 7.5 | 19 | 20.3 | 7.2 |
| H | 38 | 23.8 | 4.2 | 65 | 23.4 | 2.4 |
| I | 53 | 20.3 | 11.5 | 13 | 23.0 | 4.8 |
| All Combined | 468 | 21.5 | 8.1 | 382 | 22.3 | 5.2 |

## Percentages of Students Having Taken College Mathematics

| | University | Col.A | Col.B | Col.C | Col.D | Col.E | Col.F | Col.G | Col.H | Col.I |
|---|---|---|---|---|---|---|---|---|---|---|
| College Algebra or Pre-Calculus | 56.3 | 83.3 | 26.1 | 79.7 | 28.4 | 63.3 | 41.7 | 73.3 | 49.7 | 81.5 |
| Business Calculus | 27.7 | 58.3 | 30.4 | 56.4 | 8.1 | 30.6 | 6.9 | 20.0 | 19.0 | 0.0 |
| Mathematics of Finance | 25.8 | 56.3 | 21.7 | 59.3 | 7.4 | 14.3 | 5.6 | 16.7 | 16.3 | 0.0 |
| Freshman Calculus I | 43.6 | 27.1 | 65.2 | 28.5 | 96.6 | 10.2 | 15.3 | 10.0 | 46.3 | 18.5 |
| Freshman Calculus II | 39.0 | 16.7 | 60.9 | 21.5 | 96.6 | 8.2 | 13.9 | 10.0 | 39.5 | 7.4 |
| Freshman Calculus III | 35.8 | 10.4 | 52.2 | 20.9 | 96.6 | 4.1 | 12.5 | 0.0 | 32.7 | 3.7 |
| Sophomore Calculus I | 27.8 | 8.3 | 4.3 | 6.4 | 98.6 | 0.0 | 6.9 | 0.0 | 21.8 | 0.0 |
| Sophomore Calculus II | 27.0 | 4.2 | 0.0 | 5.8 | 98.0 | 0.0 | 6.9 | 0.0 | 21.1 | 0.0 |
| Sophomore Calculus III | 26.3 | 4.2 | 0.0 | 4.1 | 98.0 | 0.0 | 6.9 | 0.0 | 19.7 | 0.0 |

N=148

## COMP Results Profiled on Entering Freshmen
## Reference Group Norms Based on Percentiles for
## 8951 Freshmen at 46 Institutions




Subtests 1-6 and Total Score

## Percentages of Students Having Taken 2- or 3-Course Sequences in Natural Sciences

| | University | Col.A | Col.B | Col.C | Col.D | Col.E | Col.F | Col.G | Col.H | Col.I |
|---|---|---|---|---|---|---|---|---|---|---|
| Astronomy | 7.8 | 0.0 | 0.0 | 12.8 | 2.0 | 16.3 | 15.3 | 0.0 | 8.2 | 0.0 |
| Basic Engineering | 18.6 | 10.4 | 8.7 | 2.3 | 77.0 | 0.0 | 2.8 | 0.0 | 4.1 | 0.0 |
| Biology | 36.0 | 64.6 | 8.7 | 31.4 | 10.8 | 26.5 | 70.8 | 56.7 | 40.8 | 51.9 |
| Botany | 3.6 | 16.7 | 0.0 | 1.2 | 1.4 | 0.0 | 2.8 | 3.3 | 7.5 | 0.0 |
| Chemistry | 45.5 | 89.6 | 13.0 | 12.8 | 94.6 | 12.2 | 22.2 | 50.0 | 36.7 | 100.0 |
| Geological Science | 14.1 | 18.8 | 0.0 | 18.0 | 5.4 | 26.5 | 23.6 | 13.3 | 12.9 | 0.0 |
| Microbiology | 3.8 | 10.4 | 4.3 | 1.2 | 0.7 | 0.0 | 1.4 | 10.0 | 3.4 | 33.3 |
| Physics | 33.5 | 33.3 | 91.3 | 11.6 | 88.5 | 0.0 | 15.3 | 13.3 | 25.2 | 0.0 |
| Zoology | 8.5 | 8.3 | 0.0 | 1.7 | 2.7 | 0.0 | 6.9 | 20.0 | 12.9 | 74.1 |

## Percentages of Students Having Taken the Following Social/Applied Sciences

| | University | Col.A | Col.B | Col.C | Col.D | Col.E | Col.F | Col.G | Col.H | Col.I |
|---|---|---|---|---|---|---|---|---|---|---|
| Anthropology | 31.3 | 33.3 | 8.7 | 15.7 | 15.5 | 63.3 | 37.5 | 50.0 | 38.8 | 96.3 |
| Audiology & Speech Pathology | 4.1 | 4.2 | 0.0 | 1.2 | 0.0 | 4.1 | 19.4 | 0.0 | 5.4 | 3.7 |
| Economics | 71.1 | 85.4 | 17.4 | 94.2 | 82.4 | 95.9 | 40.3 | 93.3 | 49.7 | 11.1 |
| Geography | 33.8 | 37.5 | 4.3 | 36.6 | 42.6 | 32.7 | 48.6 | 16.7 | 27.9 | 0.0 |
| Human Services/ Social Work | 11.0 | 0.0 | 0.0 | 11.6 | 2.0 | 6.1 | 34.7 | 16.7 | 11.6 | 22.2 |
| Political Science | 41.3 | 27.1 | 4.3 | 47.7 | 23.6 | 98.0 | 43.1 | 26.7 | 49.7 | 18.5 |
| Psychology | 74.9 | 75.0 | 56.5 | 71.5 | 66.2 | 95.9 | 97.2 | 83.3 | 67.3 | 92.6 |
| Sociology | 61.6 | 56.3 | 21.7 | 64.0 | 48.0 | 95.9 | 79.2 | 76.7 | 52.4 | 88.9 |

## Percentages of Students Having Taken the Following Areas in a History Sequence of at Least Two Courses

| | University | Col.A | Col.B | Col.C | Col.D | Col.E | Col.F | Col.G | Col.H | Col.I |
|---|---|---|---|---|---|---|---|---|---|---|
| Afro-American History | 1.3 | 2.1 | 0.0 | 0.6 | 0.7 | 0.0 | 0.0 | 0.0 | 4.1 | 0.0 |
| American History | 36.2 | 33.3 | 26.1 | 36.0 | 21.6 | 20.4 | 58.3 | 33.3 | 52.4 | 14.8 |
| Cultural Studies | 3.4 | 2.1 | 0.0 | 1.7 | 1.4 | 2.0 | 0.0 | 6.7 | 8.8 | 7.4 |
| Western Civilization | 21.8 | 10.4 | 13.0 | 11.6 | 6.8 | 95.9 | 13.9 | 6.7 | 36.7 | 18.5 |
| World Civilization | 2.8 | 2.1 | 8.7 | 1.2 | 4.7 | 2.0 | 1.4 | 0.0 | 4.1 | 0.0 |

## Percentages of Students Having Taken the Following Areas of the Humanities

| | University | Col.A | Col.B | Col.C | Col.D | Col.E | Col.F | Col.G | Col.H | Col.I |
|---|---|---|---|---|---|---|---|---|---|---|
| Art—History or Appreciation | 22.5 | 12.5 | 39.1 | 13.4 | 15.5 | 20.4 | 40.3 | 40.0 | 32.0 | 7.4 |
| Art—Studio | 12.2 | 12.5 | 52.2 | 5.8 | 2.0 | 20.4 | 9.7 | 23.3 | 20.4 | 7.4 |
| Dance | 12.4 | 16.7 | 8.7 | 10.5 | 4.7 | 12.2 | 25.0 | 16.7 | 15.0 | 11.1 |
| Literature —Modern or Classical | 63.4 | 37.5 | 13.0 | 73.8 | 47.3 | 81.6 | 73.6 | 70.0 | 76.9 | 33.3 |
| Music —History or Appreciation | 28.6 | 14.6 | 4.3 | 19.8 | 22.3 | 46.9 | 63.9 | 40.0 | 29.3 | 22.2 |
| Music —Performance | 10.2 | 8.3 | 4.3 | 9.3 | 4.7 | 16.3 | 19.4 | 6.7 | 11.6 | 14.8 |
| Philosophy | 38.0 | 33.3 | 26.1 | 27.9 | 26.4 | 44.9 | 45.8 | 46.7 | 49.7 | 77.8 |
| Religious Studies | 22.6 | 6.3 | 8.7 | 14.5 | 16.9 | 30.6 | 33.3 | 33.3 | 33.3 | 33.3 |
| Speech & Theatre | 44.8 | 58.3 | 8.7 | 55.8 | 23.0 | 79.6 | 55.6 | 50.0 | 42.2 | 18.5 |

TABLE 2

## Student Use of General University Programs and Services: Means by College

| | MEAN RESPONSES[1] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Service | University | Col. A | Col. B | Col. C | Col. D | Col. E | Col. F | Col. G | Col. H | Col. I |
| Main Library* | 2.5 | 2.5 | 2.5 | 2.3 | 2.4− | 2.8 | 3.0 | 2.8 | 2.3 | 2.6 |
| Undergraduate Library* | 3.4 | 3.5 | 3.5 | 3.2 | 3.2− | 3.3 | 3.0 | 3.0 | 3.6 | 3.4 |
| Career Planning & Placement Center* | 1.4 | 1.3 | 1.4 | 1.4 | 1.5+ | 1.4 | 1.2 | 1.1 | 1.5 | 1.2 |
| Counseling Center* | 1.3 | 1.3 | 1.5 | 1.3 | 1.1− | 1.2 | 1.2 | 1.3 | 1.1 | 1.2 |
| Recreation or Intramural Activity* | 2.5 | 2.4 | 2.7 | 2.9 | 2.7+ | 2.2 | 2.6 | 2.4 | 2.4 | 2.4 |
| Writing Lab | 1.1 | 1.1 | 1.1 | 1.1 | 1.1 | 1.1 | 1.1 | 1.1 | 1.2 | 1.1 |
| Student Employment Services | 1.4 | 1.4 | 1.3 | 1.5 | 1.5+ | 1.5 | 1.4 | 1.4 | 1.3 | 1.4 |
| Health Services | 2.0 | 2.0 | 2.0 | 2.2 | 2.0 | 2.2 | 2.0 | 2.4 | 2.0 | 2.1 |
| Campus Plays* | 2.1 | 2.2 | 1.9 | 1.9 | 1.8− | 2.2 | 2.0 | 2.0 | 2.0 | 2.5 |
| Campus Film Series* | 2.3 | 2.4 | 2.1 | 2.2 | 2.4+ | 2.1 | 2.5 | 2.1 | 1.8 | 2.3 |
| Campus Concert Series | 2.1 | 2.2 | 1.9 | 2.1 | 1.8− | 2.2 | 2.0 | 1.9 | 2.0 | 2.1 |
| Computer Services* | 1.8 | 1.6 | 2.4 | 1.7 | 2.5++ | 1.2 | 2.4 | 1.1 | 1.2 | 1.3 |
| University Bookstore | 3.7 | 3.7 | 3.6 | 3.6 | 3.7 | 3.7 | 3.7 | 3.8 | 3.7 | 3.7 |
| N | 809 | 170 | 118 | 68 | 86 | 72 | 62 | 73 | 65 | 83 |

[1] Higher means indicate greater use
* P .05

NOTE: ++ Highest of all colleges
+ Higher than University Mean
− Lower than University Mean
0− Lowest of all colleges

TABLE 6

## Perceived Quality of General University Programs and Services: Means by College

| | MEAN RESPONSES[1] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Service | University | Col. A | Col. B | Col. C | Col. D | Col. E | Col. F | Col. G | Col. H | Col. I |
| Main Library* | 2.9 | 2.9 | 3.1 | 2.9 | 3.0+ | 2.9 | 2.8 | 3.0 | 2.8 | 2.8 |
| Undergraduate Library* | 3.0 | 2.9 | 3.1 | 3.0 | 3.0 | 3.0 | 2.7 | 3.0 | 3.3 | 2.9 |
| Career Planning & Placement Center | 2.6 | 2.5 | 2.8 | 2.6 | 2.8+ | 2.7 | 2.3 | 2.9 | 2.6 | 2.4 |
| Counseling Center | 2.5 | 2.6 | 2.4 | 2.6 | 2.7+ | 2.5 | 2.1 | 2.7 | 2.4 | 2.3 |
| Recreation or Intramural Activity | 2.9 | 2.9 | 2.9 | 3.1 | 2.9 | 3.1 | 2.9 | 3.2 | 3.0 | 2.6 |
| Writing Lab | 2.4 | 2.1 | 2.1 | 2.5 | 2.5+ | 2.7 | 1.5 | 3.5 | 2.3 | 2.4 |
| Student Employment Services | 2.5 | 2.4 | 2.6 | 2.4 | 2.7+ | 2.4 | 2.5 | 2.6 | 2.7 | 2.0 |
| Health Services | 2.6 | 2.5 | 2.8 | 2.5 | 2.6 | 2.4 | 2.5 | 2.5 | 2.8 | 2.5 |
| Campus Plays | 3.2 | 3.2 | 3.1 | 3.2 | 3.0- | 3.3 | 3.1 | 3.3 | 3.1 | 3.0 |
| Campus Film Series | 2.9 | 3.0 | 2.9 | 3.0 | 2.9 | 2.8 | 2.8 | 3.1 | 2.8 | 3.0 |
| Campus Concert Series | 2.9 | 3.0 | 2.8 | 3.0 | 2.7- | 2.8 | 2.9 | 3.0 | 2.9 | 2.7 |
| Computer Services | 2.5 | 2.5 | 2.5 | 2.7 | 2.6+ | 2.3 | 2.9 | 3.0 | 2.5 | 2.4 |
| University Bookstore | 3.0 | 2.9 | 3.1 | 2.9 | 3.0 | 3.0 | 2.6 | 3.2 | 3.3 | 2.8 |
| N | 809 | 170 | 118 | 68 | 86 | 72 | 62 | 73 | 65 | 83 |

[1] Higher means indicate higher rating of service
* P .05

TABLE 9

## Student Ratings of University Services: Means by College

| | MEAN RESPONSES[1] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Service | University | Col. A | Col. B | Col. C | Col. D | Col. E | Col. F | Col. G | Col. H | Col. I |
| Admissions* | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 2.8 | 2.7 | 2.9 | 3.0 | 2.7 |
| Registration | 2.4 | 2.4 | 2.4 | 2.5 | 2.5* | 2.3 | 2.2 | 2.4 | 2.4 | 2.2 |
| Student Records | 2.9 | 2.8 | 2.9 | 3.0 | 3.1** | 2.8 | 2.8 | 2.9 | 3.0 | 2.8 |
| Student Loans | 2.3 | 2.4 | 2.2 | 2.6 | 2.4* | 2.3 | 2.2 | 2.2 | 2.3 | 2.2 |
| Student Grants | 2.3 | 2.4 | 2.2 | 2.3 | 2.4* | 2.3 | 2.1 | 2.2 | 2.3 | 2.1 |
| Scholarships | 2.5 | 2.4 | 2.4 | 2.6 | 2.5 | 2.8 | 2.0 | 2.3 | 2.5 | 2.5 |
| Student Conduct Office | 2.5 | 2.5 | 2.4 | 2.6 | 2.6* | 2.5 | 2.1 | 2.5 | 2.3 | 2.4 |
| Treasurer's Office | 2.5 | 2.4 | 2.6 | 2.6 | 2.8** | 2.5 | 2.4 | 2.6 | 2.6 | 2.6 |
| Drop/Add Procedure | 2.2 | 2.2 | 2.1 | 2.2 | 2.3* | 2.0 | 2.2 | 2.3 | 2.3 | 2.2 |
| Preregistration | 2.9 | 2.9 | 2.9 | 3.0 | 3.0* | 3.0 | 2.6 | 3.1 | 3.1 | 2.7 |
| N | 809 | 170 | 118 | 68 | 86 | 72 | 62 | 73 | 65 | 83 |

[1] Higher means indicate higher rating
* P .05

TABLE 12

## Rating of Quality of Programs and Services in the Major: Means by College

| | MEAN RESPONSES[1] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Major Service/Program | University | Col.A | Col.B | Col.C | Col.D | Col.E | Col.F | Col.G | Col.H | Col.I |
| Availability of Advisor* | 2.8 | 2.8 | 2.8 | 3.2 | 2.5- | 3.0 | 2.7 | 2.9 | 2.7 | 2.7 |
| Willingness of Advisor to Help* | 2.9 | 3.0 | 2.9 | 3.5 | 2.7- | 3.1 | 2.7 | 3.0 | 3.0 | 3.0 |
| Quality of Printed Program Information* | 2.6 | 2.7 | 2.5 | 2.6 | 2.4- | 2.8 | 2.1 | 3.0 | 2.6 | 2.7 |
| Helpfulness of the Office Staff* | 2.8 | 2.8 | 2.8 | 3.2 | 2.8 | 2.9 | 2.4 | 3.0 | 2.8 | 2.8 |
| Quality of Special Events* | 2.6 | 2.5 | 2.6 | 2.7 | 2.5- | 2.8 | 3.0 | 2.9 | 2.6 | 2.8 |
| Adequacy of Preparation by Lower Division Courses for Upper Division Courses* | 2.8 | 2.8 | 2.8 | 2.8 | 2.8 | 2.7 | 2.4 | 3.0 | 2.7 | 2.5 |
| Quality of Courses: Providing General Educ.* | 2.9 | 3.0 | 2.9 | 3.0 | 2.8- | 3.0 | 2.6 | 3.2 | 3.1 | 2.8 |
| Quality of Courses: Preparing for Employment* | 2.8 | 2.7 | 2.8 | 2.8 | 2.8 | 2.9 | 2.6 | 3.3 | 2.9 | 2.5 |
| Availability of Required Courses for the Major* | 2.4 | 2.5 | 1.9 | 2.6 | 2.3- | 2.4 | 2.7 | 3.0 | 2.7 | 2.2 |
| Availability of Desired Courses for the Major* | 2.4 | 2.5 | 2.1 | 2.6 | 2.3- | 2.5 | 2.5 | 2.9 | 2.7 | 2.4 |
| Organization of the Curriculum* | 2.8 | 2.8 | 2.8 | 2.8 | 2.7- | 2.7 | 2.3 | 3.1 | 2.8 | 2.6 |
| Fairness of Grading* | 2.7 | 2.8 | 2.7 | 3.0 | 2.6- | 3.0 | 2.4 | 2.7 | 2.8 | 2.5 |
| Quality of Instruction in Lower Division Courses in the Major* | 2.6 | 2.7 | 2.5 | 2.7 | 2.4- | 2.7 | 2.6 | 2.8 | 3.0 | 2.6 |
| Quality of Instruction in Upper Division Courses in the Major | 3.1 | 3.0 | 3.1 | 3.2 | 3.0- | 3.0 | 3.1 | 3.3 | 3.2 | 3.0 |
| Opportunities for Interaction with Faculty in the Major* | 2.6 | 2.5 | 2.5 | 3.2 | 2.4- | 2.9 | 2.8 | 3.0 | 2.8 | 2.7 |
| Practicum/Intern in the Major* | 2.6 | 2.3 | 2.3 | 2.7 | 2.7+ | 3.2 | 2.0 | 3.1 | 3.2 | 2.9 |
| Library Collection Related to the Major* | 2.7 | 2.7 | 2.7 | 3.0 | 2.6- | 2.7 | 2.3 | 3.0 | 3.1 | 2.7 |
| N | 809 | 170 | 118 | 68 | 86 | 72 | 62 | 73 | 65 | 83 |

[1] Higher means indicate higher rating
* P .05

TABLE 15

## Rating of Quality of Facets of the Classroom Experience: Means by College[1]

| | MEAN RESPONSES[2] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | University | Col.A | Col.B | Col.C | Col.D | Col.E | Col.F | Col.G | Col.H | Col.I |
| Comprehensiveness of Course Content | 3.0 | 3.0 | 2.9 | 3.1 | 2.9- | 2.9 | 3.1 | 3.0 | 3.0 | 2.8 |
| Relevence of Content for Student Needs | 2.8 | 2.8 | 2.8 | 2.9 | 2.8 | 2.8 | 3.0 | 2.9 | 3.0 | 2.7 |
| Extent to which Content Is Current | 3.2 | 3.2 | 3.1 | 3.3 | 3.1- | 3.1 | 3.3 | 3.2 | 3.4 | 3.1 |
| Instructor's Class Presentations | 3.0 | 2.9 | 3.0 | 3.0 | 2.8- | 2.9 | 3.1 | 3.0 | 3.0 | 3.0 |
| Instructor's Class Preparation | 3.2 | 3.3 | 3.2 | 3.3 | 3.1- | 3.1 | 3.2 | 3.1 | 3.3 | 3.3 |
| Instructor's Enthusiasm for Teaching This Class | 3.3 | 3.3 | 3.2 | 3.3 | 3.2- | 3.1 | 3.4 | 3.1 | 3.3 | 3.3 |
| Instructor's Help with Problems | 3.1 | 3.1 | 3.0 | 3.0 | 3.0- | 3.0 | 3.3 | 3.0 | 3.1 | 2.9 |
| Fairness–Testing | 2.9 | 3.0 | 3.0 | 2.9 | 2.9 | 2.7 | 3.1 | 2.9 | 2.8 | 2.9 |
| Fairness–Grading | 3.0 | 3.0 | 3.0 | 3.0 | 2.9- | 2.8 | 3.1 | 2.9 | 2.9 | 2.9 |
| Clarity of Course Objectives | 2.8 | 2.8 | 2.9 | 2.8 | 2.7- | 2.8 | 2.8 | 2.9 | 2.8 | 2.7 |
| Conduciveness of Climate–Learning | 2.8 | 2.8 | 2.7 | 2.8 | 2.8 | 2.5 | 2.9 | 2.7 | 2.8 | 2.6 |
| Relevance of Lecture Information to Course Objectives | 3.1 | 3.1 | 3.0 | 3.2 | 3.1 | 3.0 | 3.1 | 3.1 | 3.1 | 3.0 |
| Quality of Classroom Discussion | 2.7 | 2.6 | 2.6 | 2.8 | 2.6- | 2.6 | 2.8 | 2.7 | 2.8 | 2.6 |
| Accuracy of Catalog Description–Course | 2.7 | 2.7 | 2.7 | 2.7 | 2.8+ | 2.7 | 2.6 | 2.8 | 2.6 | 2.7 |
| Instructor's Knowledge of Subject Matter | 3.6 | 3.6 | 3.5 | 3.6 | 3.5- | 3.4 | 3.6 | 3.6 | 3.7 | 3.5 |
| Instructor's Availability for Consultation | 3.0 | 3.1 | 3.0 | 3.0 | 2.8- | 2.9 | 2.9 | 3.0 | 3.1 | 3.0 |
| Overall Quality of Instructor | 3.2 | 3.2 | 3.2 | 3.3 | 3.0- | 3.0 | 3.4 | 3.2 | 3.2 | 3.1 |
| Overall Quality of Course | 2.9 | 2.9 | 2.9 | 3.0 | 2.8- | 2.8 | 3.0 | 2.9 | 3.0 | 2.8 |
| N | 809 | 170 | 118 | 68 | 86 | 72 | 62 | 73 | 65 | 83 |

[1] College in which course is being taken

[2] Higher means indicate greater use

* P .05

TABLE 18

## Selected Student Opinions and Characteristics: Means by College

| Characteristic | University | Col.A | Col.B | Col.C | Col.D | Col.E | Col.F | Col.G | Col.H | Col.I |
|---|---|---|---|---|---|---|---|---|---|---|
| Sex: Male | 52% | 48% | 61% | 65% | 82%++ | 12% | 80% | 8% | 23% | 35% |
| Female | 48% | 52% | 39% | 35% | 18% | 88% | 20% | 92% | 77% | 65% |
| Residence: Dormitory | 44% | 42% | 42% | 31% | 55% | 52% | 52% | 51% | 47% | 40% |
| Apartment | 27% | 25% | 29% | 33% | 28% | 28% | 35% | 26% | 20% | 27% |
| Fraternity | 3% | 5% | 4% | 9% | 4% | 0% | 4% | 0% | 1% | 2% |
| Home | 26% | 28% | 25% | 26% | 14% | 20% | 9% | 24% | 22% | 31% |
| Community of Origin: | | | | | | | | | | |
| Farm | 5% | 5% | 3% | 27% | 5% | 11% | 5% | 7% | 2% | 4% |
| Non-farm Rural | 11% | 9% | 7% | 24% | 12% | 10% | 15% | 18% | 11% | 17% |
| Town Under 2,500 | 3% | 2% | 1% | 3% | 4% | 3% | 5% | 6% | 8% | 2% |
| Town 2,500-25,000 | 16% | 13% | 21% | 4% | 19% | 19% | 11% | 14% | 17% | 13% |
| Town 25,000-100,000 | 24% | 26% | 21% | 19% | 24% | 21% | 26% | 19% | 35% | 24% |
| City Over 100,000 | 42% | 45% | 48% | 22% | 38% | 36% | 39% | 37% | 28% | 39% |
| Hours Working: | | | | | | | | | | |
| None | 54% | 53% | 50% | 46% | 65%+ | 60% | 63% | 66% | 57% | 40% |
| 1-9 Hours | 9% | 8% | 9% | 15% | 12% | 7% | 10% | 7% | 9% | 7% |
| 10-19 Hours | 15% | 18% | 13% | 18% | 9% | 15% | 18% | 13% | 14% | 22% |
| 20 or over Hours | 22% | 21% | 28% | 22% | 14%- | 18% | 10% | 13% | 20% | 31% |
| Proportion of Education Paid by Parents: | | | | | | | | | | |
| More Than Half | 59% | 62% | 68% | 52% | 48% | 67% | 63% | 58% | 57% | 63% |
| Less Than Half | 15% | 13% | 10% | 24% | 26% | 17% | 17% | 15% | 11% | 19% |
| None | 26% | 25% | 21% | 24% | 26% | 17% | 20% | 26% | 32% | 18% |
| Number of Close Relationships with Faculty: | | | | | | | | | | |
| None | 48% | 41% | 55% | 35% | 69%++ | 44% | 47% | 47% | 37% | 46% |
| One | 13% | 14% | 12% | 9% | 12% | 13% | 16% | 7% | 16% | 20% |
| Two | 17% | 23% | 10% | 21% | 10%- | 19% | 10% | 23% | 25% | 14% |
| Three or More | 22% | 23% | 23% | 35% | 10% | 24% | 27% | 23% | 22% | 20% |
| Hours on Campus Each Week Outside of Class: | | | | | | | | | | |
| Under 10 | 29% | 31% | 27% | 29% | 21%- | 22% | 7% | 38% | 45% | 34% |
| 10-20 | 14% | 14% | 14% | 19% | 17% | 17% | 13% | 10% | 9% | 12% |
| 20-30 | 8% | 7% | 9% | 13% | 4% | 14% | 11% | 8% | 5% | 11% |
| Over 30 | 49% | 47% | 50% | 38% | 59%+ | 48% | 69% | 44% | 42% | 43% |
| Fraternity or Sorority? | | | | | | | | | | |
| Yes | 22% | 22% | 35% | 29% | 13% | 26% | 13% | 28% | 12% | 28% |
| No | 78% | 78% | 65% | 71% | 87%+ | 74% | 87% | 72% | 88% | 72% |
| Satisfied with Social Experience at U.T.K.? | | | | | | | | | | |
| Very Satisfied | 36% | 36% | 41% | 41% | 25%-- | 34% | 29% | 52% | 36% | 41% |
| Somewhat Satisfied | 48% | 46% | 44% | 44% | 60%++ | 49% | 47% | 38% | 50% | 43% |
| Somewhat Dissatisfied | 13% | 14% | 12% | 10% | 13% | 16% | 19% | 6% | 9% | 16% |
| Very Dissatisfied | 4% | 4% | 3% | 4% | 2%- | 1% | 5% | 4% | 5% | 0% |
| Satisfied with Academic Experience at U.T.K.? | | | | | | | | | | |
| Very Satisfied | 20% | 15% | 24% | 16% | 17%- | 22% | 21% | 36% | 32% | 16% |
| Somewhat Satisfied | 52% | 51% | 50% | 51% | 58%+ | 56% | 55% | 49% | 40% | 46% |
| Somewhat Dissatisfied | 23% | 26% | 21% | 29% | 21% | 17% | 23% | 12% | 23% | 32% |
| Very Dissatisfied | 5% | 7% | 5% | 3% | 4%- | 6% | 2% | 3% | 5% | 6% |

APPENDIX E

# Student Satisfaction Survey Excerpt from Report for Department A

TABLE 1

## Rating of Quality of Facets of the Classroom Experience: Department A Enrolled Students

| | RATING | | | | MEANS[1] | | |
|---|---|---|---|---|---|---|---|
| | Excellent | Good | Fair | Poor | Dept. | College | University |
| Comprehensiveness of Course Content | 33% | 53% | 10% | 5% | 3.1 | 2.9 | 3.0 |
| Relevence of Content for Student Needs | 58% | 35% | 3% | 5% | 3.5 | 2.8 | 2.8 |
| Extent to which Content Is Current | 49% | 41% | 8% | 3% | 3.4 | 3.1 | 3.2 |
| Instructor's Class Presentations | 24% | 47% | 21% | 8% | 2.9 | 2.9 | 3.0 |
| Instructor's Class Preparation | 37% | 40% | 13% | 11% | 3.0 | 3.1 | 3.2 |
| Instructor's Enthusiasm for Teaching this Class | 37% | 37% | 16% | 11% | 3.0 | 3.1 | 3.3 |
| Instructor's Helpfulness with Student Problems | 40% | 37% | 18% | 5% | 3.1 | 3.0 | 3.1 |
| Fairness of Testing | 24% | 44% | 21% | 12% | 2.8 | 2.7 | 2.9 |
| Fairness of Grading | 24% | 42% | 24% | 11% | 2.8 | 2.8 | 3.0 |
| Clarity of Course Objectives | 21% | 49% | 21% | 10% | 2.8 | 2.8 | 2.8 |
| Conduciveness of Climate for Learning | 15% | 56% | 21% | 8% | 2.8 | 2.5 | 2.8 |
| Relevance of Lecture Information to Course Objectives | 36% | 49% | 6% | 9% | 3.1 | 3.0 | 3.1 |
| Quality of Classroom Discussion | 25% | 47% | 19% | 8% | 2.9 | 2.6 | 2.7 |
| Accuracy of Catalog Description–Course | 15% | 39% | 36% | 9% | 2.6 | 2.7 | 2.7 |
| Instructor's Knowledge of Subject Matter | 69% | 18% | 13% | 0% | 3.6 | 3.4 | 3.6 |
| Instructor's Availability for Consultation | 26% | 46% | 20% | 9% | 2.9 | 2.9 | 3.0 |
| Overall Quality of Instructor | 49% | 28% | 18% | 5% | 3.2 | 3.0 | 3.2 |
| Overall Quality of Course N = 43 | 37% | 47% | 13% | 3% | 3.2 | 2.8 | 2.9 |

[1] Higher means indicate higher rating

TABLE 2

## Rating of Quality of Department Programs and Services: Department A Majors

| | RATING | | | | MEANS[1] | | |
|---|---|---|---|---|---|---|---|
| Service | Excellent | Good | Fair | Poor | Dept. | College | University |
| Availability of Advisor | 23% | 40% | 26% | 11% | 2.8 | 3.0 | 2.8 |
| Willingness of Advisor to Help | 29% | 41% | 21% | 10% | 2.9 | 3.1 | 2.9 |
| Quality of Printed Program Information | 21% | 38% | 30% | 11% | 2.7 | 2.8 | 2.6 |
| Helpfulness of the Office Staff | 23% | 46% | 25% | 6% | 2.9 | 2.9 | 2.8 |
| Quality of Special Events | 16% | 47% | 28% | 9% | 2.7 | 2.8 | 2.6 |
| Adequacy of Preparation by Lower Div. Courses for Upper Div. Courses | 13% | 52% | 29% | 8% | 2.7 | 2.7 | 2.8 |
| Quality of Courses in Providing General Educ. | 19% | 55% | 23% | 3% | 2.9 | 3.0 | 2.9 |
| Quality of Courses in Preparing for Employment | 17% | 54% | 23% | 3% | 2.9 | 2.9 | 2.8 |
| Availability of Required Courses for the Major | 14% | 46% | 26% | 15% | 2.6 | 2.4 | 2.4 |
| Availability of Desired Courses for the Major | 14% | 50% | 24% | 13% | 2.7 | 2.5 | 2.4 |
| Organization of the Curriculum | 16% | 57% | 20% | 7% | 2.8 | 2.7 | 2.8 |
| Fairness of Grading | 12% | 70% | 14% | 4% | 2.9 | 3.0 | 2.7 |
| Quality of Instruction in Lower Division Courses in the Major | 10% | 54% | 29% | 7% | 2.7 | 2.7 | 2.6 |
| Quality of Instruction in Upper Division Courses in the Major | 27% | 58% | 13% | 2% | 3.1 | 3.0 | 3.1 |
| Opportunities for Interaction with Faculty in the Major | 19% | 39% | 31% | 11% | 2.7 | 2.9 | 2.6 |
| Practicum or Internship Experiences in the Major | 60% | 26% | 12% | 2% | 3.4 | 3.2 | 2.6 |
| Library Collection Related to the Major | 9% | 52% | 33% | 6% | 2.6 | 2.7 | 2.7 |

N = 75

[1] Higher means indicate higher rating

TABLE 3

## Percentage Distributions for Selected Socio-Demographic Characteristics: Department A Majors

| | Dept. | Univ. |
|---|---|---|
| Age | | |
| 18 or Under | 8% | 8% |
| 19 | 14% | 21% |
| 20 | 22% | 18% |
| 21 | 22% | 19% |
| 22 | 19% | 13% |
| 23 or Over | 16% | 22% |
| Sex | | |
| Male | 18% | 52% |
| Female | 81% | 48% |
| Marital Status | | |
| Single | 87% | 87% |
| Married | 11% | 11% |
| Divorced | 3% | 2% |
| Current Residence | | |
| Dormitory | 38% | 44% |
| Apartment | 38% | 27% |
| Fraternity | 3% | 3% |
| At Home | 22% | 26% |
| Family Income | | |
| Under $10,000 | 12% | 8% |
| $10,000 - $19,999 | 16% | 16% |
| $20,000 - $29,999 | 4% | 18% |
| $30,000 - $39,999 | 28% | 17% |
| Over $40,000 | 40% | 42% |
| Community of Origin | | |
| Farm | 11% | 5% |
| Non-farm Rural | 11% | 11% |
| Town Under 2,500 | 8% | 3% |
| Town 2,500 - 25,000 | 16% | 16% |
| Town 25,000 - 100,000 | 27% | 24% |
| City over 100,000 | 27% | 42% |
| Hours Working | | |
| None | 73% | 54% |
| 1 - 9 Hours | 3% | 9% |
| 10 - 19 Hours | 11% | 15% |
| 20 or Over Hours | 14% | 22% |
| Grade Point Average | | |
| 0 - 1.9 | 3% | 5% |
| 2.0 - 2.5 | 30% | 26% |
| 2.6 - 2.9 | 46% | 27% |
| 3.0 - 3.5 | 19% | 31% |
| Over 3.5 | 3% | 11% |
| Proportion of Education Paid for by Parents | | |
| More than Half | 70% | 59% |
| Less than Half | 5% | 15% |
| None | 24% | 26% |
| Number of Close Relationships with Faculty | | |
| None | 46% | 48% |
| One | 19% | 13% |
| Two | 19% | 17% |
| Three or More | 6% | 22% |
| Hours on Campus Each Week Outside of Classes | | |
| Under 10 | 35% | 29% |
| 10 - 20 | 14% | 14% |
| 20 - 30 | 3% | 8% |
| Over 30 | 47% | 49% |
| Fraternity or Sorority Member | | |
| Yes | 19% | 22% |
| No | 81% | 78% |
| Satisfied with Social Experience at U.T.K. | | |
| Very Satisfied | 43% | 36% |
| Somewhat Satisfied | 41% | 48% |
| Somewhat Dissatisfied | 16% | 13% |
| Very Dissatisfied | 0% | 4% |
| Satisfied with Academic Experience at U.T.K. | | |
| Very Satistied | 19% | 20% |
| Somewhat Satisfied | 51% | 52% |
| Somewhat Dissatisfied | 22% | 23% |
| Very Dissatisfied | 8% | 5% |

# Guidelines for Self-Study Document

A self-study document is prepared by the academic unit prior to the review Copies are distributed to members of the review team at least two weeks prior to the on-campus evaluation. This outline for the self-study document is designed to provide guidelines and assist the unit, not to prescribe a rigid format for its content.

## Function

Specify clearly the primary function of the academic unit, including immediate and long-range goals and/or objectives for instruction,* research, and public service, within the broader context of the college and the total university.

## Program

Provide a brief statement describing the academic program, including. role, scope, breadth, and depth. Also, describe the program and its components in terms of emphasis on preparation for teaching, research and/or region.

## Faculty

Provide a brief vita for each faculty member. Describe the faculty in terms of strengths and weaknesses, particularly as related to the role and scope of the academic program. Include information in the summary concerning the past three years for the following:

1. Special teaching, research, or professional practice awards to the faculty.
2. Publications in referred journals.
3. Books published.
4. Monographs or manuals published.
5. Journals edited or number of faculty members serving on editorial boards.

6. Grants and contracts awarded from agencies external to the University.
7. Presentations at national meetings.
8. Number of faculty approved to teach 6000 (doctoral) level courses.
9. Number of faculty approved to direct doctoral dissertations.
10. Any information about faculty quality collected from students or alumni.
11. International experience.
12. Faculty development activities.

### Students

Provide student information for the past five years. Include any information considered appropriate, but at least the following if available.

1. Number of applicants.
2. Number of students admitted.
3. Number of women, minorities and international students enrolled.
4. Number of students graduated.
5. Number of students who dropped out.
6. An assessment of the quality of students as indicated by test scores, grade-point averages, or other data.
7. Financial assistance available, including number of students awarded assistantships and fellowships.
8. Description of student recruitment procedures.

### Library

Provide an assessment of the adequacy of library holdings for the program.

### Physical Facilities

Provide a brief summary of the physical facilities and describe their effect on the academic program. Include a statement concerning any pertinent equipment needs.

### Program

1. Admission procedures—Describe how students are selected for the program.
2. Statement of desired outcomes of instruction for students.
3. Innovative, unique, or outstanding features of the program.
4. Breadth and depth of program—Include in this statement any special degree requirements, requirements for courses outside the academic unit, the selection of a student's committee, the nature of the comprehensive examination and the dissertation.
5. Research in the program—Include any information concerning how students are involved in the research, whether research assistantships

are availabie, how the research is funded, the emphasis upon research as a component of the program.

6. Public Service—Include specifically the interrelationships between public-service activities and research and other aspects of the program.
7. Teaching—Include information concerning any innovations as well as assessments by students, faculty or alumni.

## Indicators of Program Quality

1. Evidence of effectiveness of the general education component of the curriculum (for undergraduates), including value-added calculation for the college using scores on the ACT COMP exam.
2. Evidence of the achievement of desired outcomes of instruction for students, including results of comprehensive examinations and regional/national competitions, and documentation of placement and career success by graduates.
3. Assessments of program quality by enrolled students (both majors and non-majors) and graduates.

*1983-1984 Additions Underlined

National Center for Higher Education Management Systems