When I recently faced the fact that I would be called upon to make a pres­
idential address (after first wondering if the custom that brings us together
this evening is not already anachronistic) I quite naturally turned to the
printed versions of the addresses of my predecessors. John Stecklein with his
memorable lost page, Jim Montgomery with his circular administration building,
and Lee Hull, with his "fact, myth, and doubt" have more than ably discussed
the history, the present and the future of our Association and of our work,
institutional research. They could not be topped. As a consequence, I decided
to address myself to a specific substantive issue in higher education, one that
it seems to me, should be of concern to each of us. The issue is the assess­
ment of program quality.

Everyone talks about quality or excellence, but nobody does anything about
it. Now, this is an overstatement, certainly; but it is sometimes fair, I
think, to overstate in order to make a point. It has been our corporate in­
ability to come to grips in specific and meaningful ways with the concept of
quality in educational programs that underlies my contention that not very much
has been done about it. To be sure, quality has been viewed as low student­
faculty ratios and this view has guided actions. Quality has also been viewed
as distinguished faculty and some of these have been recruited in its name.
Indeed it is the latter view that seems to underlie the well-known Cartter
report. These views may be applauded, if only on the basis that they have pro­
vided grounds for action. They are at least superior to the view that quality
is intangible, undescribable, and, consequently, unassessable. The logical
implication of this view is that in fact there is nothing that a college or
university can do about the quality of its programs. If this were the case we
would do better to stop talking about it, and get on with more practical prob­
lems. But this is an area that we can not in good conscience treat in so cava­
lie a fashion. We must, I think, continue the search.

It is clear to me, at least for now, that the student-faculty ratio, dis­
tinguished faculty, and similar single-dimension views of quality in higher
education are overly restricted and may be inappropriate guides to action. I
would hope that what I have to say might not only convince you that these views
are too narrow, but would also suggest an alternative view that has some merit.
It is easy to be critical, but criticism is most helpful when alternatives are
proposed.

My remarks derive from a very intensive effort we have initiated the past
year at the University of Missouri. This effort has been directed towards
developing a concept of and assessment procedures for graduate program quality.
Consequently, I can claim little, if any, credit for any originality that may
appear in what follows; my colleagues have contributed much. I will, of course,
accept responsibility for the limitations of these remarks. Indeed, the ideas
expressed here have already received some fire in discussions with University
of Missouri faculty groups, following their review of the position papers which
we prepared as a stimulant to these discussions.
We have directed our first efforts to graduate programs and you may notice this focus in some of my remarks. We had to start somewhere and our efforts will be extended to undergraduate programs. I don't apologize for not giving you a complete and final set of procedures for the assessment of quality. It is important that in this controversial area, there be room for you, your administrative officers, and your faculties to reason together on central issues and procedures.

Purposes of Quality Assessment

Why should we attempt to assess program quality? Simply stated, because we talk about it all the time. A little more specifically, each of our colleges and universities is committed to maintaining and developing quality in its programs and in the absence of a concept and evidence of quality this commitment is hollow. Even more to the point is the fact that program quality is what education is all about and this fact should force institutions of higher education to take it into explicit consideration in all matters of decision making, management, and planning.

Another approach to answering this question of why quality assessment is needed lies in what it should be able to do for us. I will suggest five specific purposes of program quality assessment.

New Programs

The expansion of knowledge and its reorganization into new categories, the natural desire to expand offerings to higher degree levels, the attempt to attract attention through breadth of offerings, and a sincere desire to satisfy social needs for trained manpower in special areas all prompt our faculties to ask for authorization to offer new programs. New program proposals compete among themselves and with existing programs for scarce resources. On what basis should decisions for approval of such proposals be made? Clearly, in any rational process, the mission of the institution and the relation of the proposed program to this mission and to existing programs is a central consideration, as is student and social demand for the program. The most crucial consideration, however, has to be the degree of quality promised by the program. If a meaningful concept of program quality exists, then proposed programs can at least be ranked on the basis of quality promised and, other considerations being equal, the program of highest promised quality can be given priority. Even if the other considerations are not equal, no college or university with integrity will institute a new program without a promise of some acceptable level of quality, no matter how significant the need for the program on other bases. Program quality should be a central consideration in the review of new program proposals and this consideration should be based upon a meaningful concept of quality and how it may be assessed.

Existing Programs

The assessment of quality of existing programs can do two important things for the college or university that is interested in implementing its commitment to quality. First, it can provide a ranking or, at least, a categorization of
these programs on the basis of their relative quality. (As an aside I will mention here that it is probably inappropriate to compare diverse programs, say philosophy and physics, by the same standards. But in some way or another the diverse programs must be reflected on a common quality scale.) The ranking or categorization of programs may lead to the identification of some programs of high quality. These programs, in general, must be supported and nourished in a manner that will maintain this quality and the institutional leadership it signifies. Another group of programs will be identified as possessing adequate, but not high, quality. With this identification, and by bringing other obvious considerations to bear, decisions as to which of these programs should be maintained at just this level of quality and which should be nourished with the specific intent of raising their quality to the high category can be made. Finally, some programs will certainly be identified as possessing substandard quality. It is largely within this group that candidates for discontinuance can be found. If a substandard program is not marked for termination, the institution automatically marks it for specific attention directed towards increasing its quality to at least the "adequate quality" category. Specific evidence on quality may be the critical factor in permitting an institution to do what everyone knows should from time to time be done, but seldom is, and that is to prune academic programs.

The second service that quality assessment of existing programs can provide is to identify the specific program characteristics of the program which need attention if its quality is to be raised. This point leads me to the third potential benefit of quality assessment.

Guidelines for Budget Support

The idea of program quality assessment should provide guidelines for use by departments in requesting budget support for program or quality improvement purposes. The identification of program weaknesses should be expected to reveal the specific manners in which money may be deployed to alleviate the weaknesses and thereby enhance overall program quality.

To digress a moment, I would point out that the quality dimension is so central to higher education that its explicit consideration is what should distinguish program-planned budgeting for higher education from its application to other areas of human endeavor. With our computerized and even our implicit models of college or university operation, budgets for program continuation can be calculated in a largely automatic fashion on the basis of enrollment projections, course enrollment cross-over matrices, inflation factors, and the likes. Others have dealt with this topic; I will not pursue it here. The point is that the automatic calculation of continuation budgets allows the institution to focus program and budget attention where it should be focused—upon program improvement and new program budgets. Departments which identify specific weaknesses in their programs through quality assessment may, with the specificity required by the idea of program-planned budgeting, request money for program improvement. Program improvement requests then may be allowed to compete in a meaningful manner in the arena of scarce resources and, perhaps, with new program requests. Priorities can then be attached to such requests and they can be funded as far down the list as the money lasts. As with others of my statements, this one is oversimplified, but it does point up what I mean by using program quality assessment to facilitate budgeting for quality improvement.
Cost-Benefit Analysis

Quality assessment is also the basis for the cost-benefit analysis of new program and program quality improvement expenditures. By this I mean that after a new program has been instituted, the degree to which the quality it promised has developed should be determined, and ex post facto reviews of the impact of program improvement budgets should be carried out. Quality assessment is the means for these types of analyses.

Necessity of Research

Finally, and in a slightly different vein, the following proposals for quality assessment must be subjected to research in order to estimate the procedure's validity and as a basis for revisions in the procedure. In language this group will understand, what I have to propose may be considered to be a model. The research I speak of here would be directed towards a validation of this model.

A Concept of Quality for Assessment

I hope it has not been procrastination which has led me to delay so long this evening in facing up to necessity of suggesting a definition of program quality. At this point I am beginning to feel like the chef in the restaurant which advertised that 'We serve everything.' A patron, noticing the sign with this slogan, ordered an elephant sandwich. The waiter took the order and after a consultation with the chef, had to inform the customer that they couldn't kill a whole elephant for just one sandwich.

The matter of program quality is certainly an elephant. Perhaps, if I can produce a single sandwich for us to chew on without killing the elephant, that will be enough. Clearly, the quality of an educational program is indicated by the impact of the program on the behavior of the students who participate in that program. This statement has direct implications for program quality assessment and, indeed, it has served as a guide for some research on quality. For any particular program, of course, the statement needs to be fleshed out. What behaviors is the program designed to change and in what directions? Behavior, of course, must be broadly viewed and probably needs to include covert and affective as well as overt and intellective behavior. Also, how can behavior change be measured or assessed? Again, others have dealt with these questions and I will not review their work. However, I will assert that while as recently as ten years ago the measurement technology for first approximation assessment of relevant behavior changes did not exist, such tools are available to us now. We are merely avoiding the problem if we continue to hide behind the folklore that it cannot be done.

But there are problems in restricting our view of program quality to that provided by the behavior change framework. One of these is that we should be concerned with quality assessment as a basis for action and the behavior change approach to quality assessment takes time. For undergraduate programs it takes four years, at least, plus another number of months for data analysis. In any ultimate sense, product measurement would take place at some point in time after the student has left the program. In either case, programs change and the re-
sulting quality assessment would have dubious relevance to the program which existed when the evidence was available.

An even more fundamental problem is that while behavior change may be the ultimate indicator of quality, it provides at best only an indirect assessment of program quality. It cannot be expected to reveal why a program possesses the quality revealed by the assessment. It does not shed light on specific relationships between characteristics of the program and the program's overall quality and it is these characteristics that must be manipulated in an action-effort aimed at program quality. The point is that, though student behavior indicates quality, it does not provide a definition of it. The quality of a program lies in the program itself. Consequently, the assessment of quality must be based upon measures of characteristics of programs. A pseudo-definition of program quality, then, is that it is the combination of program factors that influence student behavior change. Now, I struggled for a few minutes in framing this definition and these words do not satisfy me altogether. Perhaps I would have done just as well to leave quality as an undefined term. You can be the judge.

Where this leads us, however, is to an attempt to identify those program factors or variables that do or should be expected to influence student development and to devise methods for measuring and evaluating these factors. What we are really doing is developing assumptions or hypotheses which assert that given program factors relate to quality as indicated by desirable behavior change. At this point we are ready to begin assessing program quality by means of measurement on program variables. Without forgetting the basic assumptions, we can agree, I think, that it is such a set of measures which provide direct quality assessment and the basis for action.

Furthermore, these direct program measures, in combination with behavior change measures, provide the basis for specifically examining and revising the basic assumptions or hypotheses and of modifying the quality assessment procedure to the end of increasing its validity in reference to student growth. Action on the quality front could be delayed until evidence from this research is in, but if we are impatient, and perhaps we should be, the assumptions and evidence on program variables can provide guides to action while the research evidence is being assembled.

Program Variables for Quality Assessment

What are the program variables that may be relevant to quality? The approach we have used in the University of Missouri effort in identifying measurable program variables related to quality has been based upon a view of a university as a handler of knowledge. A university carries out the interrelated activities of generating knowledge, of transmitting knowledge and those attitudes and intellectual skills and abilities associated with generating, processing, storing, and applying knowledge. Allied with these activities are the integration, analysis and synthesis of knowledge through scholarly and creative work which includes dealing with affective knowledge. This framework has led us to the specification of categories of program variables which reflect (1) the availability of existing knowledge in the field of the program, (2) the existence of effective knowledge transmittal procedures, and because
our effort focuses upon graduate programs, (3) the existence of productive activity in developing new knowledge.

For this paper I will modify these categories slightly and suggest a set of six categories that can be used as a framework for identifying quality variables. You will recognize in all of this the ideas of inputs, resources, processes, and outputs. Thus, perhaps I am talking about a model and, indeed, one day we may be able to identify the linkages in such a way that it will fully deserve to be called that. For the moment, though, I am interested in merely directing attention to the assessment of program quality and in initiating first steps. Therefore, I will not rely heavily on the model concept. The six categories I will use are (1) Program Rationale, (2) Knowledge Resources, (3) Transmission Variables, (4) Research and Scholarly Activities, (5) Support Resources, and (6) Operating Characteristics. Let me deal with each in turn.

Program Rationale

The assumption here is that a quality program will be guided by meaningful goals and objectives and that the design of such a program will be based explicitly on these objectives. Now, it is possible that an anarchial program is at least as effective as one that is guided by objectives, but, as has been amply discussed in other contexts, the nature of desired behavior change and measurement of it are incompatible with an absence of objectives.

It should be clear that my concern is with programs of instruction and with objectives for such programs. Although colleges and universities do intermingle instructional, research, and other activities and should do so on the basis of the mutually reinforcing nature of them, it is the instructional program and the objectives for it that are of concern here and these must as clearly as possible be distinguished from the objectives of research programs. While it may be possible to nourish both birds with a single feed, I would suggest that efforts to improve instruction cannot be completely equated with efforts to improve research, even at the graduate level.

Much can be said about objectives. I want to emphasize just one point and that is that objectives must indicate the competencies the program is intended to develop in its students. One issue for graduate programs is whether teachers, researchers, or teacher-researchers are being trained. Questions of this type must, we all know, be faced.

The second part of the program rationale category has to do with the design of the program, the catalog description of the courses and related experiences provided for the students in the program. It is obvious that this design must be based on the objectives of the program. If an objective of a graduate program is to prepare its students for college teaching, one would expect that provision for relevant experiences, perhaps teaching internships, would appear in the program design.

Program rational variables can be identified and measured for individual programs. At a minimum the existence of meaningful objectives and of a related program design can be scored one, and their absence, zero. Similarly, the relevance of the design to the objectives can be rated. Perhaps even specific questions that program objectives should be expected to answer can be developed.
and serve as a basis for judgement. Probably, there is no need to go much beyond this, because there is no reason why a program rationale should not be complete and completely rational.

Knowledge Resources

The assumption underlying this category of variables is simply that program quality is dependent upon the existence of the knowledge that defines the program in convenient places and in retrievable forms. Existing knowledge is the base which is transmitted to students and upon which new information is developed. This base should be not only appropriately broad, but also open-ended. It suggests the existence in the program of faculty, trained to the doctorate, in the specialties required by the program design and of retrievable library and related resources.

The library and related resource centers, either on the campus or accessible in the community or by electronic or other means, is one subcategory and faculty expertise is a second. I will not attempt to detail relevant variables in these categories. My only advice here is that in the identification of variables to reflect library resources and faculty expertise, specific attention should be devoted to the relationship of the variables to the education of students. Faculty accomplishment in research, for example, may be a mark of expertise, but one may question the degree to which it is a relevant component of knowledge resources for instruction.

The idea of a critical mass of faculty for a program will be considered under another heading, but the coverage of the program by faculty expertise is a variable that perhaps best fits here.

Transmission Variables

The assumption underlying this third category of quality variables is that effective knowledge and skill transmittal procedures contribute to student development. Perhaps it is trite to state this assumption, but this is the pattern I have chosen to follow. This category of variables may also have library or resource center and faculty components. Is there a library staff qualified to assist students in the program in an efficient and effective manner? Is available library technology used effectively for instructional purposes, etc?

The faculty transmission side obviously includes any and all of the instructional effectiveness variables that are in the literature and these are certainly central variables in the quality scheme. The matter of student advising, as an indirect component of the transmission process, may be relevant. Are students advised in the directions suggested by the program design? The adequacy of thesis guidance is a variable for graduate programs and the pervasive factors of faculty accessibility, supportiveness, and rapport with students can be considered for any program.

Faculty evaluation has often included attention to initiative and efforts at innovation in instruction and curriculum, and while the results of such innovations may appear as pluses in the instructional effectiveness complex,
the existence of instructional experimentation may carry a "halo effect" that transforms faculty dedication and enthusiasm into student motivation and enhanced learning.

Finally, students themselves may initiate and carry on non-class activities in connection with the program that result in enhanced transmission of knowledge, skills, and attitudes. Student clubs and student-sponsored seminars are illustrative. As we talk about the student's responsibility for learning, we cannot ignore indicators of this type as measures of program quality.

Research and Scholarly Activities

This class of variables applies particularly, but perhaps not solely, to graduate programs. The assumption is that the climate produced by, as well as active student participation in, research and related scholarly and creative activities enhances student learning in desirable directions. Perhaps this enhancement is most directly related to training in research competence for students in graduate programs and it is this type of rationale that most directly connects research with graduate programs. The connection provided by faculty and knowledge growth as basic to undergraduate and graduate instruction is less direct.

Specific variables that could be measured with respect to this category can be identified with respect to the following four attributes of research and scholarly activity. One, the volume of this type of activity is simply the extent to which activities of this type exist. This may be considered to be a dimension of input. Two, the volume of production from the activities is the simple, unevaluated, output dimension and would be revealed by counts of completed projects, publications, and the like. It must be recognized that there will be a time lag between the activity and measurable evidence of its output. Three, the quality and significance of the production has to do with the efficacy and style, the clarity of purpose, the appropriateness of technique, the extent of development from and articulation with existing knowledge, the clarity of communication, and even its ultimate social impact. Measures would include the "referred journal" publication count, peer ratings, and others. This attribute would clearly be significant in the assessment of quality of research programs and is probably of more significance than the preceding two in the assessment of instructional program quality.

But of even more importance may be the degree and style of student involvement. Great research programs may be operated without students or at least with a style of student involvement that can be expected to contribute only marginally to student development. The extent to which graduate students are involved in ongoing faculty research and the extent to which this involvement is of a type that should be expected to contribute to their development of research competence, in particular, can be estimated.

Support Resources

This fifth category of quality variables is based upon the assumption that physical, financial, and emotional support by the college or university for a program is necessary, if maximum student learning is to occur. The effect of
these types of support variables may be mediated through other categories of quality variables, but at least in an experimental program of quality assessment these variables deserve attention. Included here are amounts and adequacy ratings of physical facilities -- space and equipment for teaching and research; levels of institutional and, perhaps, extramural financial support for the program; secretarial, clerical, and technical support; business procedures support; support from related academic departments; and even the emotional support of the administration. Many or all of these types of variables can be measured.

Operating Characteristics

This final set of quality variables is considered here because they include some traditional measures of quality and because they may be useful as indicators of program quality even if they are not as integrally related to quality as may be the variables of the preceding categories. I will not try to offer a general definition of the term "operating characteristics" nor will I attempt to state the assumptions of the relationships of these factors to student development. In some cases this type of assumption would be awkward at best. Perhaps the consideration of these variables can be justified on the basis of tradition and for the sake of completeness. I will suggest three sub-categories of program operating characteristics.

(a) Program Magnitude. While it certainly cannot be claimed that size is a direct or even highly relevant indicator of quality, it is to be expected that a quality program will attract students and faculty. Furthermore, there is some rationale that suggests that a program cannot achieve quality in the absence of some minimum magnitude which may be expressed in terms of "critical masses" of students and faculty. This may be the point at which to consider the proportion of students in the program who are studying on a full- as opposed to a part-time basis. Perhaps each program should undergo analyses designed to reveal what these critical masses for it should be and how it serves the special needs of part-time students. But I am getting close to talking about efficiency or productivity and this comes later.

(b) Program Growth. At least up to some maximum planned size a quality program should be expected to grow in magnitude (students, faculty, degrees, and resources). Program growth may be expected to reveal something about the vitality and momentum of the program.

(c) Program Efficiency and Productivity. Efficiency and effectiveness, that is, quality, are often considered to be independent, one of the other. In fact they may be separable and it is certainly desirable to avoid confusing the two. Yet, much as reliability is a requisite for validity in measurement, efficiency may be a component of quality in educational programs. Purposive and vital programs may be expected to be efficient and productive and the former two attributes are clearly related to quality. Specific variables for measurement might include some relevant student-faculty
ratio, degrees granted per faculty member, degrees per student enrollment, and average time to complete the degree. There are others.

Program Measures and Quality Assessment

To summarize to this point, we have talked about behavior change as the ultimate criterion of program quality and have suggested, without being specific, that relevant measures of behavior change can be applied. Then we have discussed six categories of program quality variables and the assumptions that relate these to program quality. I have not extended this discussion to the identification of all of the individual variables that can be derived from this sort of analysis. In our University of Missouri effort we have identified nearly one hundred program variables using this type of analysis. Very clearly, our next step will be to examine this list very carefully to the end of reducing it to some more manageable number while making certain that each category is adequately covered. This will be a critical step and we plan to do it.

What remains is to identify the specific procedures by which measures on these quality variables may be obtained and at this point I will merely suggest some of the measurement approaches that are available. Some of the measures will, initially at least, be crude and subjectivity will be involved. Clearly, equal-interval measurement will be the exception rather than the rule. The following list illustrates available measurement procedures:

1. Regular institutional operating record systems can provide data on enrollments, faculty, degrees granted, finances, space, and other items.
2. An exit questionnaire for degree recipients can be developed to yield not only evidence on behavior change and graduate competency, but also student evaluations of relevant program variables.
3. Annual activity, accomplishment, and program evaluation reports by faculty members can be used.
4. A special, more-or-less standardized annual report for each program by the department chairman may be required.
5. For proposed new programs, the proposal document should provide the basic data required and rating scales can be developed for use in conjunction therewith.
6. A program statement, including objectives, rationale, and program design should be available for each program offered or proposed. The structure of this statement can be explicitly specified and ratings of it can be made.
7. Ratings and judgements by outside consultants will be used for some program variables. The appropriate variables can be specified and rating scales standardized.
This list is not exhaustive. As specific variables for quality assessment are selected, other measurement procedures will be suggested.

What will result will be a set of measures for each program in some more-or-less standard framework. These may be aggregated by weighting according to some initially subjective basis or may be presented in a profile format as a basis for overall program quality assessments.

The resulting measures must be interpreted by reference to some set of standards. Because of basic differences among programs, these standards cannot be uniform for all programs. Ideally, national standards based upon corresponding measures for a reference group of corresponding programs across the country would be used. For many variables, at least for the present, such explicit standards will not be available. Implicit and relative standards must then be employed.

Clearly, national standards should be used where they are available. Consultants can be asked to make their evaluations on the basis of their knowledge of national standards. For other variables it may be possible to develop fixed institutional standards based upon the nature and requirements of individual programs. Critical numbers of students and faculty for programs may be established in this manner.

For those variables which require faculty or student ratings and evaluations, absolute internal standards may be required. The rating by a faculty member or student of some aspect of a program as "excellent," "fair," or "poor" may be allowed to speak for itself without reference to any outside standard or criterion.

A Program of Program Quality Assessment

You may by now be thinking to yourself that the magnitude of what I have been proposing makes the undertaking of this type of systematic program quality assessment unfeasible, if not impossible, and, to be sure I do not underestimate the magnitude of the task. However, despite the fact that we have been meeting our classes for one or two hundred years without such a program of quality assessment, we are convinced at the University of Missouri, at least, that such a program is essential. We recognize that it won't be set in motion in a day and are talking more in terms of years. I am confident that by the end of five years we will be well underway and that by then we will be making many types of academic and fiscal decisions on the basis of evidence of program quality. I challenge you to think about it in these terms, too.

I would pose one further question. Where should the responsibility for quality assessment lie? As I see it there will be some central system for recording the evidence and some uniform sets of guidelines that will be central to the effort. However, and this is fundamental, the quality assessment program must be both developed and applied on a broad base and with faculty understanding and, hopefully, acceptance and support. To this end we are now holding faculty seminars on the topic and are both communicating the intent of the effort and soliciting ideas and suggestions for it. Faculty participation is essential.
If program quality assessment is to be valid and useful, it cannot be imposed from on high. We expect that such a program will be carried out in departments in an essentially self-study manner. Specifically, responsibility will rest with the departmental faculty for each program to present through regular channels evidence of program quality in terms of agreed-upon variables and guidelines. Each department may add to the standard set of variables others which it considers relevant to its particular programs.

Central offices, for example, the office of institutional research, may serve to coordinate the effort and to serve as resources to departments for evidence that can be most readily collected centrally. But it will be the department's responsibility to assemble and present the evidence for the necessary central evaluation.

I fully expect that eventually the data base of the university will include a regularized and validated system of accounting for and reporting evidence of program quality and that this system will be as routine and as central in university management and planning as are our present systems of accounting for academic activities, students, and finances. I invite you to share this expectation with me.