

DEPARTMENTAL ASSESSMENT

*How Some Campuses Are Effectively
Evaluating the Collective Work of Faculty*

by JON F. WERGIN and JUDI N. SWINGEN



*A Publication of the AAHE Forum on Faculty Roles & Rewards
American Association for Higher Education*

DEPARTMENTAL ASSESSMENT

*How Some Campuses Are Effectively
Evaluating the Collective Work of Faculty*

by JON F. WERGIN and JUDI N. SWINGEN

Companion volume to
*The Collaborative Department: How Five Campuses Are
Inching Toward Cultures of Collective Responsibility*

The research reported in this paper
was supported by a grant from
The Pew Charitable Trusts.

Views expressed here are
those of the authors.



DEPARTMENTAL ASSESSMENT: How Some Campuses Are Effectively Evaluating the Collective
Work of Faculty
by Jon F. Wergin and Judi N. Swingen

© 2000 American Association for Higher Education. All rights reserved. Printed in the United States of
America.

The opinions expressed in this volume are solely those of the authors and do not necessarily represent
those of the American Association for Higher Education or its members.

Additional copies of this publication or Jon Wergin's *The Collaborative Department: How Five Campuses Are
Inching Toward Cultures of Collective Responsibility* (1994) are available from:

AMERICAN ASSOCIATION FOR HIGHER EDUCATION
One Dupont Circle, Suite 360, Washington, DC 20036-1110
Ph 202/293-6440, Fax 202/293-0073
www.aahe.org/catalog

AAHE Item #FR0003

ISBN 1-56377-049-0

ABOUT THE AUTHORS

Jon F. Wergin has served as senior scholar at AAHE for its New Pathways Project and has consulted widely on topics related to faculty and program assessment. He is professor of educational studies at Virginia Commonwealth University.

Judi N. Swingen received her Ph.D. in education from Virginia Commonwealth University in 1999, and currently serves as a research associate at its Graduate School.

Contents

Methods and Procedures.....	5
Campus Evaluation Practices.....	6
<i>Program review</i>	6
<i>Outcomes assessment</i>	7
<i>Specialized accreditation</i>	8
<i>Financial accounting initiatives</i>	9
<i>Internal quality assurance</i>	10
Components of Effective Evaluation at the Departmental Level	12
Organizational and Cultural Setting	12
<i>A leadership of engagement</i>	12
<i>Engaged departments</i>	13
<i>A culture of evidence</i>	14
<i>A culture of peer collaboration and peer review</i>	14
<i>A respect for differences</i>	15
<i>Evaluation with consequence</i>	15
Evaluation Policies and Practices.....	15
<i>A clear purpose that fits the culture and mission of the institution</i>	16
<i>“Spirit of inquiry” by the central administration</i>	16
<i>Tangible administrative follow-through</i>	17
Evaluation Standards, Criteria, and Measures.....	17
Recommendations for Evaluation Practice.....	21
<i>Be proactive in discussions of “quality”</i>	21
<i>Decentralize evaluation to the maximum possible extent</i>	22
<i>Recognize that evaluation is not for amateurs</i>	23
<i>Focus not just on enhancing collaboration and teamwork but also on “organizational motivation”</i>	24
Conclusion	26
References	28
Appendix	31

Faculty members — and the departments they inhabit — are being subjected to more evaluation today than ever before. There are significant increases not only in the sheer amount of evaluation, but in the *sources* of evaluation. In addition to the now-standard methods of program review and regional accreditation, more recent mandates include “outcomes assessment,” specialized and professional accreditation, and performance budgeting, to name a few. But what is the cumulative impact of all these on faculty work? Are faculty work lives changing? If so, in useful and constructive ways? Or is evaluation instead making faculty work more onerous and bureaucratic?

In this report we assess the ways in which academic departments in U.S. colleges and universities are evaluated, and we make several recommendations for improved practice.

This work continues an earlier line of inquiry, also published by the American Association for Higher Education (AAHE), as *The Collaborative Department: How Five Campuses Are Inching Toward Cultures of Collective Responsibility* (Wergin, 1994). That publication analyzed how five institutions were working to rally their faculty around a shorter and sharper list of goals, especially at the academic department level, and then negotiating with the departments how they would be held collectively accountable. The study suggested that even institutions that had made significant progress on the first task still had found it difficult to get very far on the second. In other words, focusing the mission is one thing, but developing workable unit evaluations is another. The challenges, as that report noted, are considerable:

The ideal approach would be to evaluate departments and other academic units in ways that are not too costly or time-consuming, that respect the diversity of disciplinary missions and cultures, and that promote departmental self-reflection, all while rewarding collective accomplishments appropriate to larger school and institutional missions. (13)

But if the challenges are considerable, so are the stakes. Institutions of higher education and their faculties face enormous pressures. Public expectations of higher education have increased while public confidence has declined. It would appear — at least superficially — that many colleges and universities have “permitted an erosion of the culture of professional accountability by which [they] have traditionally assured

In this report we assess the ways in which academic departments in U.S. colleges and universities are evaluated, and we make several recommendations for improved practice.

the quality and standards of their academic programs and degrees” (Dill, 1999: 9). Skepticism has taken the form of new demands for accountability and support for alternative educational systems that promise a higher educational return for the dollar, including new proprietary institutions such as the University of Phoenix.

Faculty also face changing expectations. On the one hand, faculty face pressures from their institutions (and their colleagues) to increase their scholarly productivity, thereby raising institutional prestige and research income; on the other hand, they hear about how they need to pay more attention to all of the *other* things they do: to become more proficient with information technology, to revamp their pedagogy to reach an increasingly nontraditional student population, to become better university “citizens,” to become more engaged with community and professional service, and so on. In short, “change” has often meant “do more.”

Colleges and universities have responded by engaging in various “restructuring” activities, most of which assume that large-scale institutional change will sooner or later trickle down to departments¹ and affect faculty work. Some institutions have seen real change from these approaches (e.g., Eckel, Hill, and Green, 1998); but other institutions have not (cf., Larson, 1997), and in these cases “departmental culture” is usually fingered as the culprit. In a widely cited study, William Massy and his colleagues (1994) described patterns of “hollowed collegiality” within academic departments, characterized by faculty isolation, fragmented communication, and a reluctance to engage in the kind of truly collaborative work required to develop and maintain a coherent curriculum. Massy’s findings ring true; but given the pressures for scholarly productivity, even in smaller institutions, it is not surprising that faculty members should act this way. Faculty typically are rewarded according to standards of quality

dictated by their disciplines, not by standards specific to their institutions or departments (Fairweather, 1996). Since most faculty work alone — and are rewarded for working alone — there is little faculty investment in activities that require collective action, such as responding to institutional mandates for “accountability” or “assessment.” Quite simply, many faculty members see little relationship between these mandates and the work they do or how they are rewarded for doing it. Faculty members do not necessarily reject the ideology behind reform; they simply do not see it as relevant to what they do and how they are rewarded.

Something has to give. Recent research provides some disturbing evidence of what happens when institutions and their faculties find themselves pulled in multiple and often conflicting directions. One consequence is greater stress, especially for junior, tenure-eligible faculty, who worry about being able to “do it all” and to do it all equally well (Menges et al., 1999). What is often neglected, or “satisfied,” is atten-

Something has to give. Recent research provides some disturbing evidence of what happens when institutions and their faculties find themselves pulled in multiple and often conflicting directions.

tion to teaching: Students report spending less time on learning activities yet they receive higher grades (Kuh, 1999). Faculty report becoming increasingly “disengaged” from their teaching and investing more of their time instead on research (Zusman, 1999), and college curricula have become fragmented and lack coherence (Gaff et al., 1997). Faculty stress is especially acute in universities with large numbers of academically underprepared students (Pitts, White, and Harrison, 1999).

The premise for the study reported here is that just as colleges and universities need to focus their missions and sharpen their priorities, so do faculty members. Faculty plates are already full, and adding new responsibilities without restructuring the work itself will only lead to greater stress, especially as long as evaluation and rewards are based on a “one size fits all” approach, which, in Gene Rice’s words, leads to a “culture of competitive advantage.”

Instead, what is needed is a different kind of culture, one in which faculty members are able to focus their efforts on activities that best draw upon their own skills, talents, interests, and experience, and which allows them to negotiate with their colleagues how they might best use these strengths to contribute to the work of the department. Thus, some faculty members might put relatively more effort into research, others into teaching, still others into institutional and professional service; and these areas of emphasis could shift throughout the course of a career.

Restructuring an institution to a new culture will be possible only when institutions shift the unit of analysis from the individual to the academic department; when the work of the department is evaluated as a whole, rather than simply as the aggregate of individual faculty accomplishments; and when faculty are evaluated according to their differential contributions to the group. As long as institutions respond to pressures for accountability with various exercises in strategic planning and creative mission building without also paying attention to change in how *departments* are evaluated and rewarded, they can expect only limited success.

OUR PURPOSE

In this paper we present and discuss the results of our survey on the evaluation of academic departments. Our original purpose was straightforward: to search for evaluation policies and practices that encourage constructive change in departments and a stronger culture of collective responsibility. We wanted to see (1) how these models worked, (2) what seemed most critical to their success, and (3) how key ideas might be applied to other settings.

But as the study progressed, our purpose shifted from a search for “models” to something slightly different: to identify elements of effective evaluation practice, and to put them together in a way that might provide harried administrators and faculty with a useful framework, or at least a visible point of reference.

After a brief methods section, we describe current practices, then abstract from these to describe key components of effective departmental evaluation. We end with several specific recommendations.

Methods and Procedures

We began this project in August 1998 by first reviewing the literature, both published and fugitive, including conference proceedings; posting messages on such listservs as the Association for the Study of Higher Education (ASHE) and the American Educational Research Association (AERA); arranging, through AAHE, for a mass mailing to all campus provosts; and calling upon personal networks of informants. We deliberately cast a wide net. The letter to provosts, for example, said,

[We are] looking for exemplary models, tools, and processes for evaluation of academic departments as units. [We] intend to profile “best practices” and publicize good ideas — and hence to help the [Pew Charitable Trusts] identify future opportunities for wise investment. If your institution is engaging in innovative work on departmental or program review, [we] want to hear from you.

These efforts netted information on about 130 institutions across the Carnegie categories (see Table 1 in the Appendix). For some, the information we have is sparse: letters, email messages, or brief notes from telephone conversations. Other institutions provided more extensive guidelines and reports. As we combed through these data, we decided to analyze institutional practices in departmental assessment at three levels of intensity: (1) simply noting current practice; (2) writing a short profile based on campus documents and telephone interviews; and (3) undertaking more in-depth case studies with site visits. Selection of institutions for the last category was based upon complexity of mission (no baccalaureate or two-year institutions), distinctiveness of method, comprehensiveness of approach, and, in a couple of instances, hunches that a personal visit would be worthwhile.

We visited eight institutions: Georgia State University, Indiana University of Pennsylvania, Indiana University Purdue University Indianapolis, Northwestern University, St. Mary’s University (TX), Southeast Missouri State University, the University of Arizona, and the University of Southern California.

Data-collection protocols varied with each visit and were determined by the nature of the idea or process under study. In most cases, we interviewed the provost or vice president for academic affairs, other members of the central academic administration, selected deans and chairs, and key faculty leaders. Visits lasted an average of one and a half days. We shared draft cases with institutional informants, and we incorporated corrections of fact — and sometimes inference — into our final drafts.

Campus Evaluation Practices

We were struck by the sheer amount of evaluation going on in academic departments. The collective work of faculty is evaluated in as many as five different ways, and some campuses employ all five. The problem is that often these methods are disconnected from one another.

Program review

Models for internal “program reviews” have been around for at least 15 years, and are now in place in most institutions. All but a handful of the institutions in our database employ formal program review in some form. The review either is cyclical (e.g., every five to seven years) or is triggered by financial or enrollment concerns. Typically, the department evaluates its strengths and weaknesses through a self-study and presents a plan for improvement, an external review panel visits the campus and writes a report, and a “program review committee” monitors the process.

The credibility of program reviews has suffered over the years for two reasons. First, even though program reviews are usually billed as “formative,” their impacts on departments are generally modest (Mets, 1998). A second and related problem is that most program reviews are one-shot affairs, not integrated into the life of the institution. The process often unfolds in a way that encourages participants to get through it with a minimum of aggravation. The whole thing becomes tedious, time consuming, and often of little consequence. Because the focus is backward (on what has already happened) rather than forward (on what is possible), the review becomes a ritual. The opportunity for critical reflection — a chance to put the academic values of systematic inquiry to use — is lost in the desire to get the thing done.

We did find several exceptions to this pattern, however:

- ▶ Northwestern University is a prime example of an institution where program review is taken seriously: It is set in a culture that not only values collaboration across disciplinary lines but also places a high premium on program quality. The process at Northwestern University was designed and initiated by faculty members themselves; and while a substantial infrastructure has been built to administer program review, it is clearly there to serve the faculty.
- ▶ Another example of successful program review is Georgia State University, where reviews are guided by the institution’s strategic goals and result in action plans with direct budgetary consequences.
- ▶ The University of Scranton has adopted “focused program review,” an idea borrowed from Middle States Association guidelines. Instead of undergoing the standard comprehensive review, the department may opt, with the approval of the dean, to concentrate on a shorter list of questions of the department’s own

choosing. Reports from the institution are highly encouraging, suggesting that when departments are given the flexibility, they will ask themselves the difficult questions.

Outcomes assessment

Assessment has also become a common fixture in universities, but somewhat more recently, dating back to the late 1980s. The principal drivers for institutional assessment programs have been mostly external — various state agencies and legislatures and regional and professional accreditation associations. They have shifted the focus of assessment from “inputs” (such as campus resources) to documented student learning. Unlike program review, which focuses on the academic department, institutional assessment programs cut across the educational mission at various levels: the departmental major, general education, the baccalaureate degree. Outcomes assessment has met with mixed success in most institutions. As with program review, it seems that more attention is given to the mechanics associated with collecting and reporting data, and less to determining what kinds of data should be collected and how they can be used to improve student learning. In only a few institutions does the trajectory of assessment seem to cross that of program review; in these institutions, St. Mary’s University, for example, outcomes assessment has supplanted program review by focusing assessment at the departmental level.

It seems that more attention is given to the mechanics associated with collecting and reporting data, and less to determining what kinds of data should be collected and how they can be used to improve student learning.

The two methods of evaluation do, however, share a rather dubious distinction: Faculty see the questions driving most assessment efforts as “theirs,” not as “ours.” The review is on someone else’s agenda — higher administration, governing board, professional or disciplinary society. Most faculty accept the necessity of program review and outcomes assessment, but don’t generally see these as processes that will affect their own professional practice, at least not in a positive way. Further, at most institutions, outcomes assessment data are not tied directly to the evaluation of departments as units. But as with program review, our survey turned up some notable exceptions:

- St. Mary’s University not only puts assessment at the center of institutional priorities, but plans are mainly department based. While the administration makes it clear that assessment results will be important determinants of resource allocation at the school and departmental levels, individual units are encouraged to develop assessment plans that inform internal curricular decisions. Further, each department at St. Mary’s identifies a faculty member to coordinate assessment and to help colleagues interpret the data.

- Worcester Polytechnic Institute has for the past 25 years required students to take three capstone courses — including one in the humanities — that require an integrative project. Student products not only are used to certify student competence but also serve collectively as the basis of annual departmental self-evaluations of teaching effectiveness.
- Ohio University, a research institution, also has a successful history of outcomes assessment at the departmental level. It has used “department-based assessment” since 1995; each year departmental faculty are asked to prepare a brief report focusing on evidence of completion of student learning objectives, both for majors and for other students taking their courses. Departmental reports are forwarded to the dean for review, and then on to the provost. An issue at Ohio is a tendency on the part of faculty to view assessment as top-down; as one informant noted, perhaps the best way to shift the focus to one of program improvement is to involve faculty members closely in developing a new process.
- At Baruch College, part of the City University of New York (CUNY) system, this problem has been met head-on with a “faculty-centered approach” to departmental review. Here — as with the other institutions mentioned — annual reviews are focused on learning outcomes, but departments are encouraged to be creative within general college and school guidelines.

Specialized accreditation

William Dill and others have written extensively about the limitations of specialized accreditation: What began as a mechanism for quality control in medical education has “metastasized” into nearly 100 specialized accrediting bodies, each holding local programs to standards that often ignore an institution’s distinctive mission and goals (Dill, 1998). However, increasing numbers of specialized and professional accreditation agencies have rewritten their standards to focus on how well a program meets its learning goals *in ways consistent with institutional mission*, rather than on rigid, nationally normed standards.

The most striking of these are the guidelines proposed by the Accreditation Board for Engineering and Technology (ABET) and the AACSB—The International Association for Management Education. ABET, for example, has developed Engineering Criteria 2000, an approach that purports to increase the flexibility of accreditation criteria by focusing on standards of three general types: the degree to which the school practices “continuous quality improvement”; the degree to which students have the knowledge required for *entry* into the engineering profession; and the degree to which the school provides support *adequate for the program’s objectives* (emphasis added). In this way, said Kate Aberle, associate executive director for ABET,

Several things happen. First, institutions and programs define their mission and objectives to meet the needs of their constituents, and thus enable program differentiation. Second, our emphasis on outcomes extends to basic preparation for professional practice. And third, we leave it up to the programs to demonstrate how these criteria and their educational objectives are being met.” (personal communication, September 1998)

Other accrediting bodies now looking to similar, school-based criteria include the American Dental Association, the American Physical Therapy Association, the American Psychological Association, the American Council on Pharmaceutical Education, the National League for Nursing, and the American Occupational Therapy Association.

Financial accounting initiatives

Several universities have developed an accounting procedure called “activity-based costing” (ABC), which has been adapted from the business community. The principle underlying this method is “subsidiarity,” or the idea that decisions about allocating resources need to be made at the level responsible for implementing those decisions. Two higher education models that employ modified techniques of ABC are the Stanford Cost Model (or “contribution margin analysis”) and responsibility-centered management (RCM).

Contribution margin analysis subtracts costs from revenues for each school or other standard “cost unit.” That figure represents the funds available to support the infrastructure of that unit (the “contribution margin”). A spreadsheet is then used to determine each unit’s ratio of revenue to infrastructure cost. Since Stanford University operates on a central budget, the Stanford Cost Model has not been used for allocation decisions. It was designed instead to be one part of a programmatic review that would include such other factors as contributions to teaching, research productivity, national quality rankings, and so on. Nevertheless, the model has been used at Stanford University to help justify the reorganization of several departments.

Responsibility-centered management picks up where contribution margin analysis leaves off, by decentralizing fiscal responsibility and authority and pushing decision making to the school or department level. RCM also makes units responsible for generating income and managing expenses, and allows savings to be carried forward.

In theory, the key advantages of RCM are that (1) units develop increased awareness and accountability, (2) they gain flexibility and control over how funds are used, and thus (3) they are more motivated to improve program quality. Our sense, based on our review of RCM at the University of Southern California and Indiana University Purdue University Indianapolis, is that only the first two steps are achieved. That is, units do in fact become more fiscally responsible, and they do put their flexibility to creative use. We did not, however, see much evidence of greater at-

tention to program quality under RCM, unless “quality” is defined as maximizing profit or minimizing loss. On the other hand, a common complaint about RCM — that it “balkanizes” academic units and thus reduces incentives for cooperation — seems overstated.

Internal quality assurance

Each of the above forms of unit evaluation is essentially “top-down” or “outside-in”: that is, the stimulus is external to the department, and the emphasis is on accountability to external constituencies. But unit evaluation focused solely on accountability is incomplete. Accepting the need to be “accountable” is one thing; developing a more internalized sense of *responsibility* for quality is another. Research on personal motivation has shown that while external incentives may be important to gain our attention, *internal* motivation is what sustains us day-by-day (cf., Csikszentmihalyi, 1997).

While this promotion of CQI is still in the formative stages, it is an example of how institutions might truly restructure faculty work in ways that promote collaboration and collective responsibility.

Faculty members thrive to the extent that they gain intrinsic satisfaction from the work they do. But how are faculty to feel connected to an intellectual community in an increasingly privatized world? How, more particularly, might departmental faculty develop a sense of collective responsibility for their work, a sense that their individual effectiveness is inseparable from the effectiveness of the group?

Our survey turned up several examples of institutions that promote “continuous quality improvement” (CQI) at the departmental level; but few seem able to make the idea work. One moderately positive example is Indiana University of Pennsylvania’s Reflective Practice Project. At IUP, departmental teaching circles have evolved into deeper departmental conversations about expectations for students, how these expectations translate into performance indicators, and how these indicators in turn relate to institutional measures. Seattle Pacific University is exploring ways of linking a series of institution-wide initiatives with “departmental development plans,” which in turn are linked to individual faculty workplans. While this promotion of CQI is still in the formative stages, it is an example of how institutions might truly restructure faculty work in ways that promote collaboration and collective responsibility. Finally, Northwest Missouri is one of the few institutions that has focused its discussions about “quality” at the level of the academic department. In the words of provost Tim Gilmour,

We learned that we have to talk with each department individually: “How do we work together to find out what we need to know to get better? . . . We can’t tell you what the best measures are — but there has to be a design and there have to

be measures that tell you and us how well you're doing." (personal communication, June 1999)

But these cases are exceptional. In general, campus practices of quality assurance at the departmental level often suffer from two debilitating problems: lack of relevance and little coordination.

First, most departments and most faculty do not see the relevance of such practices to the work they do. The notion of "continuous quality improvement" has not taken hold. Faculty are already so busy with research and teaching that the idea of using evaluation for formative purposes is lost. Further, faculty view institutional measures of "quality" as off the mark, as not congruent with what their own definitions of quality might be. Consequently, program review and outcomes assessment exercises often have only marginal impact.

By contrast, successful campuses have a common theme: Faculty and departments have a strong influence on the purposes, processes, and methods of evaluation and on the evaluation questions asked. From this we infer that at the institutional level, demands for unit accountability should focus less on accountability for achieving certain predetermined results and more on how well units conduct evaluations for themselves and use the data these evaluations generate. This notion is similar to David Dill's ideas about "academic audit" (1999) and William Massy's ideas about "quality-process review" (1997): Rather than attempting to evaluate quality itself, the focus instead is on processes believed to *produce* quality. These ideas have sprung mostly from work in Western Europe (particularly the United Kingdom, Denmark, and the Netherlands) and Hong Kong, and they are directed mostly to the assessment of entire institutions; of those countries, only Denmark has a tradition of using the audit as a means of evaluating individual academic units (Thune, 1999).² Audits have yet to emerge as a viable alternative model in the United States at any level.

Second, there is little coordination at most institutions (except perhaps for scheduling) among assessment, program review, and external accreditation. It is ironic that the premium placed on administrative efficiency has not yet extended to activities that are presumed to make the institution more efficient. We were unable to locate a single institution where these activities complemented and informed one another well, with the possible exception of IUPUI, where all reviews of professional programs are negotiated to add value to their accreditation processes. Ohio University, among others, is working on improving coordination.

One additional problem should be noted. Even institutions with effective, workable systems of program review, such as Northwestern University, show huge variations across departments and schools. Sometimes campus policies seem to make little difference; what does matter is effective unit leadership, at both the school and departmental levels. We address this issue next, as we lay out a preliminary framework describing the qualities of effective departmental assessment.

Components of Effective Evaluation at the Departmental Level

What makes evaluation of academic departments “effective”? We would suggest that effective evaluation informs judgments of quality that then lead to improved departmental functioning. Two key points are embedded in this statement: that evaluation informs judgments, it does not dictate them; and that evaluation promotes constructive change.

Our research to date suggests that when defined in this way, effective departmental assessment depends on three key factors:

- ▶ the degree to which the *organizational and cultural setting* promotes a conducive atmosphere for evaluation;
- ▶ the credibility and fairness of *evaluation policies and practices*; and
- ▶ the validity and reliability of *evaluation standards, criteria, and measures*.

We describe each factor in turn below.

ORGANIZATIONAL AND CULTURAL SETTING

The most important first step to quality assurance is not finding the right instrument or technique, but rather building an institutional climate supportive of quality improvement. When Northwestern University provost Lawrence Dumas was asked how he would go about initiating program review in another institution, he said this: “First I’d take a measure of the institution and its vision for the future. Is there ambition for change? I would try to find ways of articulating a higher degree of aspiration; if there weren’t a strong appetite for this then program review would be doomed to failure” (personal communication, January 1999).

Here are some elements of a “quality” institutional climate as suggested by the institutions we reviewed.

A leadership of engagement

This is admittedly an all-encompassing term, but it characterizes leaders who are able to frame issues clearly, put clear choices before the faculty, and be open to negotiation about what will inform these decisions. In short, these are leaders who are able to make a clear and compelling case for why change is in the interest of the school or department and what the consequences of inaction will be. Leaders who engaged the faculty in this way were able to avoid the sort of “compliance mentality” that has