

FACULTY EVALUATION

*By Professor Michael Theall – Section Editor -- Director CATALYST
Youngstown State University*

The **NTLF** resource library is a cooperative project undertaken in cooperation with the POD Network. The POD Network is the oldest professional organization in the United States devoted to professional, faculty, and organizational development within higher education. To learn more about POD and its dedicated efforts to enhance the quality of college teaching visit <http://www.podnetwork.org>

Faculty evaluation is a term used to describe a range of processes and practices for determining work performance of academics, usually driven by the need for decision-making one of two areas - formative (improvement) or summative (judgmental). The literature on faculty evaluation is quite extensive, encompassing over 4,000 works on topics such as student ratings of instruction (the most commonly collected data used in faculty evaluation), peer/colleague evaluation, the construction and use of performance portfolios, psychological measurement, and personnel process and policy. A significant amount of research on faculty evaluation is closely associated with faculty development studies and initiatives.

Faculty have traditionally been held responsible for (and thus rewarded based upon) their performance in three areas; teaching, scholarship, and service. A particular emphasis on any of these three areas is usually tied to an institution's mission. For example, faculty at small liberal arts colleges are often valued more for their teaching effectiveness than their research productivity, while the reverse may hold true at a research university. Historically, most types of institutions - from colonial colleges to land grant universities - rewarded their faculty for effective teaching and service, but early in the twentieth century the universities began to move toward a European model of higher education that considered research to be the primary responsibility of the institution's faculty. Dramatic increases in federal funding for research during the Cold War - providing a new stream of income for both faculty and their universities - fueled a shift at these institutions toward an emphasis on evaluating scholarly productivity. This emphasis on research and publication assumed that good scholars were good teachers, but over time students, parents, and the general public began to demand evidence of teaching effectiveness as well, and by 1980, student ratings of instruction were common throughout higher education. Major contributions to the literature were made by Centra (1979), Doyle (1975), Scriven (1967) and others. However, it was not until the last fifteen years of the century that attempts were made to develop a broader framework for faculty evaluation that included all the roles of the faculty. Shulman (1986) described three kinds of teacher knowledge: content knowledge [knowledge of one's subject]; pedagogical content knowledge [knowledge for teaching]; and curricular knowledge [a deep blending of the other two characterized by a broad repertoire of techniques, alternatives, materials, and the ability to relate the subject to a wide range of other fields].

In 1990, Ernest Boyer suggested four categories of scholarship: discovery, integration, application, and teaching. "Discovery" refers to the traditional scholarship of investigation

and dissemination of research results to peers. Faculty who succeed in this realm are seen as creating new knowledge, and often receive grants to support their research and publication or presentation of their findings. “Integration” describes the kind of scholarship that synthesizes and gives meaning to isolated facts by putting them into perspective. Faculty demonstrating this form of scholarship go beyond strict disciplinary boundaries to bring together diverse viewpoints and to look at new questions and pressing problems. “Application” refers to a form of scholarship that uses traditional scholarship to address significant problems in order to help individuals, institutions, and society. Boyer's thesis incorporated the traditional notion of service into this aspect of scholarship. Finally, the “scholarship of teaching” describes the work of a professor becoming meaningful only when it is understood by others, and is something that both educates and entices future scholars. Boyer's reconceptualization was hailed as one that would allow the recognition of teaching, and the more complete and appropriate evaluation of faculty performance.

Following this reconsideration of scholarship itself, new guidelines for assessing scholarship were proposed by researchers and policymakers in the last several years. Boyer's colleagues (Glassick, Huber, & Maeroff, 1996) proposed schema for assessing scholarship and outlined the qualities of a scholar. Considerable dialogue on the “roles and rewards” of academic life, and the criteria used to evaluate the full range of faculty performance, has been led most vocally by Eugene Rice and an initiative sponsored by the American Association for Higher Education. (See: <http://www.aahe.org/initiatives/facultyroles.htm> for more information.)

Recent trends in the research literature have attempted to meld evaluation into an ongoing faculty development process, rather than viewing faculty evaluation as a stand-alone activity whose only result is a judgment on (and not a learning opportunity for) the individual faculty member. Arreola, Theall, & Aleamoni (2003) presented an expanded view of faculty work, distinguishing between the “base profession” (one's disciplinary area of expertise) and the larger “meta-profession” (the work of the professoriate as defined to include base profession expertise, teaching, and many other related areas of responsibility). The model is based on the premise that appropriate evaluation of faculty performance should begin with a full definition of the responsibilities of the individual, followed by the development of mechanisms to assess performance in the various areas, and coupled with sufficient resources for training and development of the necessary skills.

Evaluation by one's peers is common at virtually all institutions of higher education. This form of faculty evaluation is influenced considerably by academic disciplines and departments (as opposed to institutionally standardized forms of course evaluations), and thus encompasses several dimensions of departmental and disciplinary politics, collegiality, and the perspectives of one's departmental peers regarding the rigor of one's academic research. Often, a professor's research interests are so specific that nobody else in his or her department is abreast of the field, which makes a proper evaluation of research quality impossible. In these cases, a department chair or others within their department will seek other members of the academic discipline with similar interests and familiarity of the subject under review. However, the core activities of faculty evaluation are by necessity local constructs. Clearly, it makes sense for a review of research quality to be provided by someone with appropriate expertise, usually from the same or a closely allied discipline. Above all, senior

administrators or institutional governing boards participation in these faculty evaluation activities are typically met with fervent resistance from the faculty. Although recent decades have witnessed a dramatic increase of attention paid to questions of institutional accountability, new approaches to faculty evaluation are still rare at most colleges and universities.

Conflicting views about student ratings of teaching

However, the actual processes of collecting, analyzing, interpreting, and using evaluation data on most campuses remain largely unaffected by this dialogue. Student ratings are the primary and in some cases, the sole source of data for summative decision making. In part, this limited practice had led to ongoing, and sometimes contentious debate about the technical and applied merits of ratings. A long history of ratings research consistently has shown these data to be reliable, valid, and useful. Cohen's (1981) often-cited multisection meta-analysis established a significant relationship between ratings and learning. Marsh's (1987) review considered a wide array of variables and concluded that ratings were reliable and valid. The string of substantial studies by Feldman (e.g., 1998) explored a number of variables in depth arriving at the basic conclusion that ratings were not significantly biased by these variables. However, a small number of studies have reported results interpreted by critics to prove ratings to be unreliable and invalid (e.g., Naftulin, Ware, & Donnelly, 1973), and these studies, along with a few newer ones, continue to be trumpeted as evidence that ratings should be dismissed or discontinued despite the fact that the positive evidence is overwhelming. The real problem is poor practice and misuse or misinterpretation of ratings data: a situation that has been decried for over a decade (e.g., Theall & Franklin, 1990)

Recommendations

The following are recommendations for good practice in faculty evaluation distilled from research and development efforts over the past four decades

- Establish the purpose of the evaluation and the uses and users of data beforehand.
- Include all stakeholders in decisions about evaluation process and policy.
- Keep a balance between individual and institutional needs in mind.
- Publicly present clear information about the evaluation criteria, process, and procedures.
- Establish a legally defensible evaluation process and a system for grievances.
- Include resources for improvement and support of teaching and teachers.
- Build a real "system" for evaluation, not a haphazard and unsystematic process.
- Establish clear lines of responsibility/reporting for those who administer the system.
- Invest in the evaluation system and evaluate it regularly.
- Use, adapt, or develop instrumentation suited to institutional/individual needs.
- Collect and analyze multiple sources of information from several situations.
- Keep ratings data and validate the instruments used.
- Produce reports that can be easily and accurately understood.
- Educate the users of ratings results to avoid misuse and misinterpretation.
- Keep formative evaluation confidential and separate from summative decision making.
- In summative decisions, compare teachers on the basis of data from similar situations.

- Consider the appropriate use of evaluation data for assessment and other purposes.
- Seek expert, outside assistance when necessary or appropriate.

New approaches require the investment of time, energy, and resources, as well as the commitment of all concerned parties. There are no quick or simple solutions, but many approaches hold the promise of allowing the development of fair, reliable, valid, accurate, credible, and accepted processes that will lead to improved evaluation. Lastly, the most promising institutional approaches to faculty evaluation - and which have the most potential to improving faculty performance - are those most closely tied to well-supported, long-term faculty development efforts.

References

- Arreola, R. A., Theall, M. & Aleamoni, L. M. (2003) Beyond scholarship. Recognizing the multiple roles of the professoriate. Paper presented at the 84th annual meeting of the American Educational research Association. Chicago: April 22. Available on-line at: <http://cedanet.com/meta>.
- Boyer, E. L. (1990) *Scholarship reconsidered*. San Francisco: Jossey Bass.
- Centra, J. A. (1979) *Determining faculty effectiveness: assessing teaching, research, and service for personnel decisions and improvement*. San Francisco: Jossey Bass.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51, 281-309
- Doyle, K. O. (1975). *Student Evaluation of Instruction*. Lexington, MA: D. C. Heath.
- Feldman, K. A. (1998) "Reflections on the effective study of college teaching and student ratings: one continuing quest and two unresolved issues." In J. C. Smart (ed.) *Higher education: handbook of theory and research*. New York: Agathon Press._
- Glassick, C. E., Huber, M. T. & Maeroff, G. I. (1997) *Scholarship assessed*. San Francisco: Jossey Bass.
- Marsh, H. W. (1987). Students evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388.
- Naftulin, D. H., Ware, J. E., & Donnelly, F. A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, 48, 630-635.
- Shulman, L. S. (1986) "Those who understand: knowledge growth in teaching." *Educational Researcher*, 15, 4-14.
- Theall, M., & Franklin, J. L. (1990) (Eds.) Student ratings of instruction: Issues for improving practice. *New Directions for Teaching and Learning* #43. San Francisco: Jossey Bass.

Annotated readings:

- Angelo, T. A., & Cross, K. P. (1993) *Classroom assessment techniques: a handbook for college teachers*. (2nd ed.) San Francisco: Jossey Bass.
- This widely read text is derived from work at the National Center for Research into Postsecondary Teaching and Learning at the University of Michigan. The book offers extensive description and application of various strategies for gathering and using

information to assess student learning. Since evidence of student learning is an important part of the larger evaluation of faculty performance, knowledge and use of effective assessment techniques is critical to good evaluation practice.

Arreola, R. A. (2000) (2nd ed.) *Developing a comprehensive faculty evaluation system*. Bolton, MA: Anker Publishing Company.

This is the best practical guide to developing and implementing a comprehensive evaluation system. It contains a thorough overview, grounded discussion of operational issues, samples of instruments and process, and proven techniques for generating dialogue and consensus. If one could have only one book to guide faculty evaluation, this should be it.

Arreola, R. A., Theall, M. & Aleamoni, L. M. (2003) Beyond scholarship. Recognizing the multiple roles of the professoriate. Paper presented at the 84th annual meeting of the American Educational Research Association. Chicago: April 22. Available on-line at: <http://cedanet.com/meta>.

This paper and associated website allow readers to access the full presentation as well as several matrices which describe in some detail, the 'meta-profession' concept and its application to faculty work. A separate section is included demonstrating how 'the scholarship of teaching and learning' represents a special case of the meta-professional construct.

Birnbaum, R. (1988) *How colleges work: the cybernetics of academic organization and leadership*. San Francisco: Jossey Bass

_____ (1992) *How academic leadership works. Understanding success and failure in the college presidency*. San Francisco: Jossey Bass.

_____ (2001) *Management fads in higher education. Where they come from, what they do, and why they fail*. San Francisco: Jossey Bass.

Birnbaum's work is clear, insightful, realistic, and very readable. His books provide many descriptive examples and the work is grounded in a substantial and broad literature base. Few authors have described higher education as clearly and realistically as Birnbaum. While this work centers on the ways higher education works, it is important to evaluation because it provides a better understanding of the context in which evaluation takes place. Evaluation systems must be responsive to the environments in which they exist, thus this literature is relevant and useful.

Bloom, S. (1998) Structure and ideology in medical education: an analysis of resistance to change. *Journal of Health and Social Behavior*, 29, 294-306.

This is an interesting view of the change process and a sobering description of "reform without change", that is, the futility of responding to the need to alter or improve something without implementing any long-lasting or meaningful changes. In conjunction with Birnbaum's text on "fads", it alerts readers to the need for making change relevant and connecting to stakeholder frames of reference. By its very nature, evaluation is often threatening, and it is potentially an explosive issue. Faculty evaluation is often (unfortunately) done in a 'pro forma' manner in response to pressure for reform or change and Bloom's article points out the dangers and frustrations associated with this error.

Boyer, E. L. (1990) *Scholarship reconsidered: the priorities of the professoriate*. San Francisco: Jossey Bass.

This short work has had tremendous impact in the decade since its publication, generating dialogue and action on campuses of all types. It's reconceptualization of scholarship, particularly the introduction of "the scholarship of teaching" as an equal partner with discovery, integration, and application has led to numerous projects and attempts to broaden the roles of faculty. Faculty evaluation must include clear and public definition of its expectations, process, and intended uses. By further defining the "priorities of the professoriate, Boyer opened a dialogue that has brought more attention to teaching and the scholarship that underlies it.

Braskamp, L. A. & Ory, J. C. (1994). *Assessing faculty work*. San Francisco: Jossey-Bass.

Centra, J. A. (1993). *Reflective faculty evaluation: enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass.

These are two of the most substantial, recent books on the general topic of faculty evaluation, both by widely respected authors.. They are comprehensive and discuss a range of issues, more theoretical than the Arreola text, but excellent additions to a basic library on the topic.

Chickering, A. W. & Gamson, Z.F. (1987) Seven principles for good practice in undergraduate

education." *Wingspread Journal*, 9 (2), special insert.

The seven principles have been used as guidelines for assessment in traditional and even on-line instruction and there has been widespread use of the faculty, student, and institutional inventories devised to measure implementation of the principles on campuses. There is debate about the extent to which these general guidelines should be used in evaluating faculty performance, but their impact can not be overlooked.

Cross, K. P. & Steadman, M. H. (1996) *Classroom research: implementing the scholarship of teaching*. San Francisco: Jossey Bass.

This work expands on Cross' work in classroom assessment to incorporate Boyer's definitions of scholarship and to provide teachers with strategies for conducting classroom research that supports better understanding of classroom process and simultaneously has the rigor to be incorporated into more traditional outlets for scholarship. Classroom research can be legitimate scholarship and its results can lead to improved teaching and learning. As such, the efforts of faculty who undertake classroom research should be considered valid evidence of professional activity and/or scholarship and should be among the considerations in evaluating performance.

Diamond, R. M. (2002) *A field guide to academic leadership*. San Francisco: Jossey Bass

This new book considers a broad range of topics related to effective leadership in higher education including a section that deals with assessment, faculty evaluation, program review, and the faculty reward system.

Farmer, D. W. (1999) Institutional improvement and motivated faculty: a case study. In M. Theall (Ed.) "Motivation from within: approaches for encouraging faculty and students to excel" *New Directions for Teaching and Learning* # 78. San Francisco: Jossey Bass.

_____ (1988) *Enhancing student learning: emphasizing essential competencies in academic*

programs. Wilkes Barre, PA: Kings College Press.

Farmer's work as an academic administrator and leader in change processes is summarized in the New Directions chapter and his specific efforts at Kings College are discussed in detail in

the book. Both are useful for those in a position to effect change through leadership. While the thrust of Farmer's work was in assessment, the principles he discusses also apply to institutional programs for evaluation.

Feldman, K. A. (1976 through 1998; see extended bibliography) A series of research syntheses in

Research in higher Education.

In a series of over a dozen articles, Feldman has provided the most in-depth reviews of specific issues relating to faculty evaluation and student ratings of instruction. These definitive works have explored all of the major issues raised as potential biases to faculty evaluation. Feldman's findings concur with other major work on evaluation, noting that in general, student ratings provide valid, reliable, and useful information.

Feldman, K. A. & Paulsen, M. B. (1998) (2nd ed.) *Teaching and learning in the college classroom*. Needham Heights, MA: Simon & Schuster Custom Publishing.

This is the most comprehensive collection of readings on this topic. It is broad and substantial, and includes landmark items like Boyer's definitions of the four kinds of scholarship, Chickering & Gamson's "Seven Principles", and also incorporates some of Feldman's studies of college teaching and Paulsen & Feldman's work on creating a campus teaching culture. This is probably the best single compendium of the research on college teaching. Its collection of articles make clear, the complexity and multi-dimensionality of college teaching and the associated notion that the evaluation of that teaching is equally complex.

Franklin, J. & Theall, M. (1989) "Who reads ratings: knowledge, attitudes, and practices of users of student ratings of instruction." Paper presented at the 70th annual meeting of the American Educational Research Association. San Francisco: March 31. ERIC # ED 306-241. This was the first large-scale study of faculty and administrator knowledge of and attitudes about ratings. It included a survey determining user knowledge of evaluation practice, basic statistics, and the literature of the field. Faculty and administrators in general scored significantly lower than staff in centers for teaching, learning, assessment, and evaluation, who, in turn, scored lower than a group of 40 experts in the field. Less knowledge of the literature and practice in the field was significantly correlated with negative views about evaluation and about students as providers of data.

Glassick, C. E., Huber, M. T., & Maeroff, G. I. (1997) *Scholarship assessed: evaluation of the professoriate*. San Francisco: Jossey Bass.

This work continues where Boyer's text left off, providing discussion and guidelines for evaluating and documenting scholarly work, particularly in the realm of the scholarship of teaching. Completed after Boyer's death, it nonetheless includes a prologue by Boyer. The outline of criteria for assessing scholarship has been widely accepted but its limitation with respect to faculty evaluation is that it considers only scholarship while thorough faculty evaluation must consider all aspects of performance.

Marsh, H. W. (1987) Students evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388.

This review of the faculty evaluation literature remains the most cited work of its kind. Its depth and breadth are unusual and its conclusions, supported by virtually all major researchers in evaluation, are that student ratings are, “1) multidimensional; 2) reliable and stable; 3) primarily a function of the instructor who teaches a course rather than the course that is taught; 4) relatively valid against a variety of indicators of effective teaching; 5) relatively unaffected by a variety of variables hypothesized as potential biases; and 6) seen to be useful by faculty...by students...and by administrators.”.

Miller, R. I. (1987) *Evaluating faculty for promotion and tenure*. San Francisco: Jossey Bass.

This text lists ten useful characteristics of effective systems that are the basis for many of the recommendations in this chapter, and includes legal issues, administrative roles, and discussion of promotion and tenure procedures.

New Directions for Teaching and Learning / Institutional Research

These series from Jossey Bass include several relevant titles. Those dealing with evaluation and closely related issues are identified below.

- Aleamoni, L. M. (Ed.) (1987) *Techniques for evaluating and improving instruction. New Directions for Teaching and Learning #31*. San Francisco: Jossey Bass.

This is the first in a series of volumes exploring evaluation issues and it includes contributions by many of the leading writers in the field.

- Anderson, R. S., Bauer, J. F. & Speck, B. W. (Eds.) (2002) “Assessment strategies for the on-line class: from theory to practice.” *New Directions for Teaching and Learning #91*. San Francisco: Jossey Bass.

On-line teaching and learning present new questions and issues in terms of the effect of context on instructional effectiveness. There are important questions about the evaluation of on-line teaching that have not yet been addressed and this volume’s emphasis on assessment can contribute to more accurate and effective evaluation of on-line teaching.

- Angelo, T. A. (1998) “Classroom assessment and research: an update on uses, approaches, and research findings.” *New Directions for Teaching and Learning #75*. San Francisco: Jossey Bass.

In its discussions of the use of assessment to improve teaching and learning; to help new instructors; and to support program accountability, this volume is a model partner to effective evaluation. Since student learning is one measure of effective teaching, assessment can and should be part of good evaluation practice,

- Kanpper, C. & Cranton, P. (2001) “Fresh Approaches to the evaluation of teaching. *New Directions for Teaching and Learning #88*. San Francisco: Jossey Bass.

This volume considers evaluation from several perspectives, connecting it to the scholarship of teaching, teaching awards, the use of technology, portfolio development, accreditation, and other issues.

- Lewis, Karron (Ed.) (2000) “Techniques and strategies for interpreting student evaluations.” *New Directions for Teaching and Learning #87*. San Francisco: Jossey Bass.

This volume deals with useful techniques for avoiding common errors in reporting and interpreting ratings results. Since such errors are one of the most common problems with ratings, the guidelines in this volume can be particularly useful.

- Ryan, K E. (Ed.) (2000) "Evaluating teaching in higher education: a vision for the future." *New Directions for Teaching and Learning* #83, Fall. San Francisco: Jossey Bass.

This volume considers past practice and new approaches, recommending practical strategies for improving current methods, and for developing new ones.

- Theall, M. & Franklin, J. (1990a) "Student ratings of instruction: Issues for improving practice." *New Directions for Teaching and Learning* #43. San Francisco: Jossey Bass

This is one of the first discussions of ratings and evaluation to focus on context and the systematic

nature of evaluation and the requirements it imposes. Several aspects of evaluation from ethical issues

to disciplinary differences are presented.

- Theall, M. & Franklin, J. (Eds.) (1991) "Effective practices for improving teaching." *New Directions for Teaching and Learning* #48. San Francisco: Jossey Bass.

This volume deals with teaching improvement but connects such efforts to gathering, analyzing, and interpreting evaluation and assessment data for purposes of improvement.

- Theall, M, Abrami, P. A. & Mets, L. (Eds.), (2001) "The student ratings debate. Are they valid? How can we best use them?" *New Directions for Institutional Research* No. 109. San Francisco: Jossey Bass.

This volume looks ahead to possible changes in practice affected by issues such as the increasing demand for distance and on-line learning, and the effects of new definitions of validity.

Pescolozodi, B. & Aminzade, R. (1999) *The social worlds of higher education: handbook for teaching in a new century*. Thousand Oaks, CA: Pine Forge Press.

This is a very interesting collection of works by sociologists and considers a range of higher education issues. Its treatment of higher education as a sociological microcosm is unique and worthwhile, and like the work of Birnbaum, it offers insights into the context for evaluation.

Scriven, M. (1967) Methodology of evaluation. In R. Tyler, R. Gagne, & M. Scriven (Eds.) *Perspectives of curriculum evaluation*. Chicago: Rand McNally & Co.

If there is a seminal article on modern evaluation, this is it. Scriven's terminology and approach are the basis for contemporary evaluation and assessment.

Extended bibliography of related publications:

The following bibliography is not complete. There are far too many items on evaluation to include here. Those works listed below are almost all in accessible publications or books. Presentations at conferences not available via ERIC or on-line, are not included. Some items connect evaluation with faculty or instructional development, institutional policy, leadership, and related topics, and are thus included. NOTE: The bibliographic list is not included here.