

Exploring The Reliability, Validity, And Utility Of A Higher Education Faculty Review Process

Kelly D. Bradley, University of Kentucky, USA
James W. Bradley, Bluegrass Community and Technical College, USA

ABSTRACT

Institutions of learning face significant calls for accountability, leading to an increased interest in measuring faculty performance. The purpose of this research is to describe and analyze the process of faculty evaluation in a college of education at a southeastern university. The process is outlined with attention given to the reliability, validity, and utility through documenting strengths and weaknesses. Included is a detailed review of the instruments utilized in the process. Suggestions for improvements and modifications are central to the outcomes, including plans for aligning the process with relevant research. The degree to which the faculty review process is deemed credible, professional, and relevant may serve as an indicator of the link between theory and practice. Potential ramifications of a weak or bias system, as faculty promotion and tenure are intertwined with this process, are of utmost importance. Legal issues, including challenges, rise to the forefront. This topic is of importance to students, faculty, and administrators, as all involved and impacted by the review process, either directly or indirectly. The applicability of this work beyond the example institution is also addressed.

Keywords: Evaluation, Faculty, Reliability, Validity, Higher Education

INTRODUCTION

A memorandum from the Provost's office of the example institution states, "The evaluation of faculty performance is one of the most important functions of department chairs, division directors and other evaluators. The purpose of the Faculty Performance Review is to provide guidance for continuing and meaningful faculty development".

A statement of importance, mixed with a dose of unoperationalized superlatives and good bit of vagueness. That being said, the intent that the faculty review process is to play a key role in the development of faculty is clear. Thus, there is a need to explore the reliability, validity, and utility of the process and the instruments and stakeholders involved in said process.

THEORETICAL FOUNDATION

This work is anchored in three general areas of consideration: (1) the importance of reliability and validity, as described by Latham & Wexley (1994), within the framework of performance appraisal; (2) the concept of utility as it relates to practical economic concerns described by Yunker & Marlin (1984); and (3) the alignment and disconnect between theory and practice.

Faculty evaluation, as a matter of institutional policy, began to take root during the early 1970's (Gardner, 1977) & (Baird, 1974). As scientific progress increasingly grew to enhance research funding, economic conditions developed that created a scenario where practical utility becomes remarkably interchangeable with economic utility. During this transition and in a market based economy, universities were eager to protect enrollment dollars

(Lindsey, 1081 & 1982). The crescendo of accountability begins with evaluations of faculty based upon issues of practical, economic, and institutional utility as described by (Yunker & Marlin, 1984). Realizing critical monetary concerns; institutions adapted existing evaluation processes to include instruments intended to quantify time spent by faculty engaged in various responsibilities of employment and the quality of how they conduct those activities. The underlying assumption being that quality academics are manifest in efficient and determined effort.

Evaluation should be anchored in reliable and valid measures. Per Latham & Wexley (1994), *validity* is achieved if an instrument measures the intended attribute; *reliability* speaks to the consistency of results across multiple data sets, or in the case of faculty performance reviews – across and between faculty, department, and colleges. Even in the absence, or great difficulty, of the ability to assign statistical significance, utilitarian concerns should not be dismissed. Fairweather (2002) suggests that “research accomplishment is a cosmopolitan academic function” (p. 26). Research accomplishments thus become increasingly equated to monetary gain. His work questions the very notion of feasible efficiency, and casts doubt on agreements of utility.

As institutional policy outlines, the importance of providing ongoing performance evaluation of faculty, the evaluation and measurement framework points to the importance of providing a similar, continual review of the process itself to capture opportunities for improvement. There is established support in the research stressing the importance of developing valid and reliable measurement tools. A quick survey of literature related to assessment, design, development, evaluation, or implementation practices generates much data describing the importance of engaging in *praxis of reliability and validity*. The complex nature of utilitarian factors is described well by Lindsay (1981 & 1982), Michalak (1981), and Ewell (1991). Each document, especially the selected instruments, has significant challenges associated with accounting for faculty productivity, utility, and less tangible activities. Recent trends, as illustrated by Rasch (2004), Reckase (2001), and Slevin (2001), raise the stakes to a new level as faculty become dependent on self-evaluation procedures requiring demonstrable effective use of time as a critical resource.

PURPOSE

Institutions of learning face significant calls for accountability. Stakeholders, including students, educators, and administration, are expected to demonstrate their effectiveness in learning, teaching, and management. The process of faculty review should support these endeavors. The extent, to which faculty review procedures are aligned, useful, and relevant, provide a proxy measure of the ability to integrate theory and practice. Here, the faculty review process is reviewed with a focus toward the reliability, validity, and utility of the practices involved. Specific attention is given to the instruments utilized to measure and record a rating for the faculty performance. The degree to which the faculty review process is deemed credible and professional may serve as an indicator of the link between theory and practice. The purpose of this research is to describe and analyze the process of faculty evaluation in the College of Education at a southeastern university. Suggestions for improvements and modifications are central to the outcomes.

An Example of the Process

The faculty performance review process is being reviewed in general, but with a specific example and attention given to the measures being used. Here, a faculty performance review at a southeastern university, within the college of education provides the framework for discussion. Outlining the process, annually, university policy requires that all tenure-track and newly hired faculty regardless of title series be evaluated. Tenured faculty are reviewed every other year, but any tenured faculty may choose to be reviewed during their off-year. The following persons are not evaluated: tenured faculty who will resign or retire before, or at the end of the current fiscal year, non-tenured faculty whose appointments will not extend beyond the end of the current fiscal year, and faculty appointments that provide no more than twenty percent of their regular salary.

Each individual being reviewed is required to provide

1. Faculty Performance Review
2. Faculty Activities Report
3. Teaching Portfolio
4. Distribution of Effort (DOE) form

These documents will be outlined below in instrumentation. In addition to the aforementioned, the faculty member may choose to provide an Optional Narrative. This can be used to document, expand, or comment on the information provided for the review. It is limited to two pages. All materials are submitted to the department chairperson. The process, inclusive of reviews and appeals, runs from January through May each year.

The following domains form the assessable areas for the performance review: (1) teaching and advising; (2) research, creative, and scholarly activities; (3) university and public service, including community engagement; and (4) administration. Reviews are to be based on the DOE performed by the faculty member. Quantitative and qualitative information are to be used in judgment about performance. The department chair reviews each faculty member, and is expected to consult with a suitable committee of peers. Then, the chair recommends an evaluation rating to the dean. The final rating is assigned by the dean.

METHOD

The research design consists of a review and critique of the documents and procedures utilized by the College of Education at a southeastern university. During the review, relevant research relating to faculty evaluation was considered, with a focus on the alignment the intentions, procedures, measures, and outcomes. Areas of concern and suggestions for improvement are offered, with the direction to give attention to each activity and relative weightings.

Instrumentation

The evaluation instruments or forms are to be developed by the dean of the college and must involve consultation with an appropriate faculty group. Evaluations must contain sufficient written commentary to explain the assigned ratings, especially if there is an unsatisfactory rating or if the need for significant improvement is cited.

Faculty Performance Review

A summary instrument, known as the Faculty Performance Review, rates faculty in each of the four domains, described previously, with weighted scoring based on the DOE, or a composite DOE if variable during the review process. While letter, numerical, or descriptive designations may be used, the rankings must clearly recognize at least three designations: outstanding, good or satisfactory, and unsatisfactory performance. Within the college of education, there are five ratings: 1–unsatisfactory, 2–marginal, 3–professional, 4–high professional, and 5–outstanding. This form provides the final quantitative scores and also has an area for the chair to comment on each area and a corresponding area for the faculty member to respond to those comments. This is followed by the chair’s signature area and then the dean’s statement to agree or not agree with the ratings along with a signature area. Finally, faculty members are to sign the faculty performance review, once the ratings have been assigned. The faculty member also makes the choice to: agree or disagree with the rating, and supply rationale where needed. When disagreeing, they may choose to simply record their disagreement, submit additional materials for review, or request a formal appeal.

Faculty Activities Report

Each faculty member is asked to document his or her effectiveness and progress for either the year or two-years under review. The faculty activities report offers a structured outline to the faculty member to detail: (1) teaching and advising; (2) research and scholarly activities; (3) university and public service; and (4) administration. Each individual under review is asked to include the products of collaboration and multi-disciplinary efforts in teaching, research, and service under the appropriate section. Teaching and advising asks for course information, advising, and dissertation committee work. Research and scholarly activities include refereed journal articles, book, book chapter, and instructional materials published or accepted for publication; papers presented at professional meetings; other writings; grants submitted; new and continuing grants received; and work in progress. University and public service requests professional organizations, university service, college and department service, and school service. Finally, administration outlines university, college and department. Each section has a final category of other for faculty to include anything else they feel is relevant evidence.

Teaching Portfolio

The teaching portfolio is composed of materials related to teaching and advising collected and maintained by the faculty member. It is intended to serve as an instrument for review, evaluation, and improvement of teaching and advising. The portfolio must contain: a reflective statement regarding teaching and advising; a list of courses for semesters under review including title, description, and number of students enrolled; representative course syllabi; a quantitative and qualitative summary of student evaluations; the number and level of undergraduate and graduate advisees and a list of masters and doctoral student committees; a list of directed thesis or dissertations; a summary of service associated with student organization and student-faculty committee service; and student evaluation of advising.

Distribution of Effort

The determination-of-effort (*DOE*) tool is intended to allow faculty an opportunity to account for expenditures of time. The categories are codified and standardized in order to construct uniform ratings as part any faculty member's overall evaluation score. Faculty are required to account for expenditures of time, as a percentage of all working hours, into four categories: (1) Teaching & Advising; (2) Research; (3) Creative & Scholarly Activities and University & Public service; and (4) Administration & Professional Development. Within each of the four categories are specific examples of qualifying activities. The activities are presented to be valid measures of the tasks research faculty are responsible for completing; including specific publication efforts as required for tenure, but not specifically identified as so.

Stakeholders

During this review process, many individuals could be considered stakeholders. To operationalize in this framework, a stakeholder is a person, group, organization, or system who affects or can be affected by the review process. The faculty member under review is an obvious stakeholder, considering such components as the time to compile the materials and the impact of the outcome of the review. There are also numerous individuals involved in the review process, such as the chair, dean, and in some circumstances teams of senior faculty members. The department chair has the time constraints of conducting the reviews and the personal and political pressures of being the lead of a merit review process. The dean wears multiple hats, as a colleague and leader, all while being responsible for equity in the process. All evaluators are required to be both fair and constructive. Beyond those directly involved, students are also stakeholders, as results could directly impact their educational environment.

FINDINGS AND DISCUSSION

Validity of the tools and process are recorded to be established through the correlation of expected faculty attributes with measured DOEs. Reliability has been confirmed with consistent faculty responses including similarly calculated scores across the College of Education. Statistically speaking, reliability is the consistency of a set of measurements or measuring instrument. Reliability does not imply validity. So, a reliable measure could be measuring consistently, but not measuring what it is suppose to measure. Here in lies the first issue, what is it that the faculty performance review is measuring? Consistent faculty attributes, expected DOE percentages... the consistent nature of the process might make it the most flawed. Formalizing the process, the tools, and the procedures is necessary, but it also allows the reproduction of materials and even replication of materials across individuals, resulting in a potentially invalid process. A logically valid argument is one where the conclusion follows from the premises. Is the premise to document activity, thus, the faculty activities report, teaching portfolio, or is it to further develop faculty utilizing these sources? Clearly, the tools and directive statements are in some disagreement.

Assessing the college's 'Distribution of Effort' instrument offered an authentic exercise in faculty review activities and provides useful questions for future study. Data is lacking on the potential for developmental power as a result of the distribution of effort exercise. Research opportunities exist for the statistical correlation of effort as an input and/or output. It may be important to determine the importance of faculty perceptions of review process utility. An important first step will be the construction of a survey instrument to gauge faculty responsiveness to the review

process and *distribution of effort*. The mathematics, time commitments, and disposition create conflict when attempting to complete a DOE. Everything must add to 100%, because theoretically, at 100% an individual should be saturated. Again, an issue of standardization becomes a conflict. Each class must count as 12.5% effort, thus if a faculty member teaches a typical 2-2 load, 50% of the DOE is consumed. Questions arise, such as, is it reasonable to assign half of a faculty member's effort to teaching, with research and service not yet accounted for, and potentially administration as well? What would it mean to record over 100% effort? Is the DOE a relative perception or is it actual behavior? The DOE is a tool that could be manipulated in this process. The faculty performance review involves weighted scoring, so if a faculty member is strong in an area and weak in another, the individual has the ability to load their DOE in this manner to create a scenario for a good review, even with a weak area.

The final outcome of the university-evaluation-process can directly impact salary and tenure determinations. A significant body of literature offers findings that support the importance of alignment and utility. Colleges of Education, in particular, require students to complete course work in assessment and evaluation. Yet, they do not always follow the guidelines that are introduced, rooted in theoretical findings. It is necessary for faculty and college administrators to 'walk the walk' and 'practice what they teach, and in some cases preach'. Even more important are the potential ramifications of a weak or bias system, as faculty promotion and tenure are intertwined with this process. Legal issues, including challenges of the review scores, rise to the forefront, especially in light of the administration structure.

The power structure of the process is yet another criticism. Faculty input is encouraged with faculty response indicators leading to a mutually agreed state of evaluation and placement. This is the intention, but this also has potential ramifications. For a faculty member to disagree with the ratings could be viewed as disrespectful, not being a team player, or even combative. The process is a power structure, and with salary, professional level, and tenure being involved, or at least connected to the performance review, the voice faculty are given is really in principal, as many untenured faculty just sign and say ok, as that is expected. On the other end of it, the scores are assigned by the chair, regardless of rank. This can put a chair, especially one that is not a full professor, in a difficult situation, such as evaluating more senior level colleagues. On paper this should not matter, but in practice, it does... at least to some. If this is true, fairness and equity do not occur.

This paper does not examine the contribution of student evaluation forms on the faculty review process, as an entire body of literature has criticized and questioned the validity and nature of student input. If the purpose of a faculty performance review truly was formative, no one would object that such input could be useful; however, given the summative, final nature of the process, giving too much weight to student evaluations as evidence of teaching effectiveness again demonstrates a potential flaw in the process.

Utility is a measure of the relative satisfaction from, or desirability of, consumption of various goods and services. Given this measure, some would argue that utility is in the eyes of the beholder for the faculty review process. Reviewing the statement from the provost's office, the goal is for this to be a meaningful process for all parties involved. The issue is satisfaction. The reflective nature of the review process has a faculty member review an entire year, or even two, of effort and relative tasks. If the review is not positive, what is the utility? Is the process really viewed as a way to improve on deficiencies and build on strengths? Or instead, is it viewed as a merit review process, where monetary outcomes are the only significant measure of utility?

Matters of assessment, measurement, and justification are imbedded in the history of public P-20 education (Wiggins, 1993). From the Normal Schools as an accreditation processes, to the correlation of funding with proficiency outcomes, and to include the critical need for research dollars in Higher Education. There exists a fascinating set of structural processes, market dependent, on the outcomes-of-evaluation. Faculty and administration tend to see themselves in an individual role in the faculty review process (Baker, 1994). The tendency to position oneself in one role often yields blindness to the others. As a result, the individuals within the system become stuck in a pattern which inhibits an ability to rationally explore alternative solutions, approaches, or measures.

EDUCATIONAL SIGNIFICANCE

Institutions of learning face significant calls for accountability. Stakeholders, including students, educators, and administration, are expected to demonstrate their effectiveness in learning, teaching, and management. The process of faculty review should support these endeavors. The extent to which faculty review procedures are aligned, useful, and relevant, serves as a proxy measure of the ability to integrate theory and practice. There is an ongoing need to maximize the effectiveness of faculty review procedures at this and other institutions. It is not clear that the outcomes are sufficient for the time and energy put into the process, especially in years that monetary awards, in the case of this specific institution – merit raises, are nonexistent. In an era where calls for salary caps and the end of tenure are the norm, it is crucial to obtain reliable and valid measures, and to have a process that all stakeholders can recognize the utility of conducting.

REFERENCES

1. Baird, L.L. (1974). The practical utility of measures of college environments. *Review of Educational Research*, 44(3), 307-329.
2. Baker, E.L. (1994). Researchers and assessment policy development: A cautionary tale. *American Journal of Education*, 102(4), 450-477.
3. Ewell, P.T. (1991). To capture the ineffable: New forms of assessment in higher education. *Review of Research in Education*, 17, 75-125.
4. Fairweather, J.S. (2002). The mythologies of faculty productivity: Implications for institutional policy and decision making. *The Journal of Higher Education*, 73(1), 26-48.
5. Gardner, D. (1977). Five evaluation frameworks: Implications for decision making in higher education. *The Journal of Higher Education*, 48(5), 571-593.
6. Latham, G.P. & Wexley, K.N. (1994). *Increasing productivity through performance appraisal*. Reading, Mass: Addison-Wesley.
7. Lindsay, A. (1981). Assessing institutional performance in higher education: A managerial perspective. *Higher Education*, 10(6), 687-706.
8. Lindsay, A. (1982). Institutional performance in higher education: The efficiency dimension. *Review of Educational Research*, 52(2), 175-199.
9. Michalak, S.J, Jr., & Friedrich, R. J. (1981). Research productivity and teaching effectiveness at a small liberal arts college. *The Journal of Higher Education*, 52(6), 578-597.
10. Rasch, L. (2004). Employee performance appraisal and the 95/5 rule. *Community College Journal of Research & Practice*, 28(5), 407-414.
11. Reckase, M.D., Feuer, M.J., & Haertel, E.H. (2001). The controversy over the national assessment governing board standards. *Brookings Papers on Education Policy*, (4), 231-265.
12. Slevin, J.F. (2001). Engaging intellectual work: The faculty's role in assessment. *College English*, 63(3), 288-305.
13. Wiggins, G. (1993). Assessment: Authenticity, context, and validity. *The Phi Delta Kappan*, 75(3), 200-214.
14. Yunker, J.A., & Marlin, J.W., Jr. (1984). Performance evaluation of college and university faculty: An economic perspective. *Educational Administration Quarterly*, 20(1), 9-37.