

## **Faculty Teaching Performance: Perceptions of a Multi-Source Method for Evaluation (MME)**

**Adrian R. Lyde,<sup>1</sup> David C. Grieshaber<sup>2</sup>, and George Byrns<sup>3</sup>**

*Abstract: Evaluating college and university faculty teaching performance is necessary for multiple reasons, including assurance of student learning and informing administrative decision-making. A holistic system of evaluating university teaching is necessary for reasons including the limitations of student evaluations and the complexity of assessing teaching performance. University faculty members were interviewed to determine their perceptions of the multi-source method of evaluating (MME) teaching performance after a revision of policies and procedures was approved. The MME is comprised of three primary data sources: student evaluations, instructor reflections describing attributes of their own teaching such as the teaching philosophy, and a formative external review. While the faculty perceived the MME as a useful tool, they still believe it operates more to produce a summative product than work as a formative process. According to the results, a more formative process would be supported by addressing several factors, including timing of reflections, accountability from year to year, and mentoring. Improving these constraints may make the proposed MME a more appropriate tool for formative review of teaching.*

*Keywords: college, university teaching evaluation, teaching effectiveness, formative evaluation, teaching development*

Measuring the quality and accountability of teaching effectiveness in higher education has a lengthy and well-researched history (Arreola, 2000; Costin, Greenwald & Menges, 1971; Clinton, 1930). Still, the questions of what “effective” means and how it is measured continue to challenge college and university faculty and administrators, particularly in regard to personnel decisions (Arreola, 2000; Sproule, 2000; McKeachie, 1997). Student ratings of instruction are the most commonly used measure of teaching effectiveness (McKeachie, 1997; Gustad, 1961). However, teaching effectiveness as a measurable construct is more complex (Young & Shaw 1999; McKeachie, 1997). What is the standard and who sets it? How is it measured objectively? How are the measurement results used? Fiscal constraints and the desire for better student outcomes contribute to increasing demands for accountability of student learning, thereby increasing the importance of evaluating university teaching effectiveness (McCarthy, Niederjohn & Bosack, 2011; Arreola, 2000). When considered holistically, teaching effectiveness can account for teaching skill and student learning, as well as the process of improving both.

---

<sup>1</sup> Department of Health Sciences, Illinois State University, Campus Box 5220, Normal, IL, 61790-5220, arlyde@illinoisstate.edu.

<sup>2</sup> Department of Health Sciences, Illinois State University, Campus Box 5220, Normal, IL, 61790-5220, dcgries@ilstu.edu.

<sup>3</sup> Department of Health Sciences, Illinois State University, Campus Box 5220, Normal, IL, 61790-5220, gebyrns@ilstu.edu.

Two forms of assessment are used to evaluate teaching for different purposes. Summative assessment is often used to judge teaching performance that impacts personnel decisions such as the awarding of tenure or promotion, but may not be helpful to the instructor (Van Note Chism, 1999; Raths & Preskill, 1982). Alternatively, formative assessment assists the instructor by providing information about teaching strengths and areas of improvement (Raths & Preskill, 1982). A clear definition of formative assessment is "...a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes..." (North Carolina Department of Public Instruction, 2008). By this definition, formative assessment is not an evaluation apart from teaching but rather an integral part of the teaching and learning process.

Given the wide range of research data related to student evaluation of college and university teaching, including how the data is interpreted, many authors have suggested that a logical approach for assessing the multi-faceted nature of teaching is to use multiple data sources in faculty evaluation (McKeachie, 1997; Cashin, 1990; Marsh, 1984). Producing multiple data sources supports faculty to develop a purposeful repository of evidence that demonstrate teaching effectiveness. To continue support for the development of new faculty member teaching effectiveness and to improve upon the skills of experienced faculty members, the evaluation criteria at a large Midwestern university was amended and the policies and procedures utilized to evaluate teaching performance were clarified according to best practices in the literature. Such clarification, the faculty believed, would support formative development of teaching while continuing to produce a summative score suitable for personnel decisions. The result of these changes was a multi-source method for evaluation (MME) comprised of three primary data sources: student evaluations, an instructor portfolio, and reflection on formative external reviews.

After two years of implementation, researchers wanted to understand faculty perceptions of the amended teaching performance evaluation system. What follows provides an overview of the MME structure, including performance evaluation levels for each data source (see Table 1). The next section provides supporting literature and best practices. The study methods, results and discussion of faculty perceptions are also presented. The proposed MME may be adaptable to assessment activities at other universities based on contextual factors that are unique to those institutions.

### **Structural Overview of Multi-Source Method for Evaluation (MME) of Teaching Performance**

The MME is comprised of three primary data sources: 1) student evaluations; 2) a portfolio prepared by faculty describing attributes of their own teaching, including, reflection on student evaluation data, development of a teaching philosophy, and construction of a professional development plan; and 3) reflection on a formative external review. The proposed MME has as its primary purpose to "facilitate growth and professional development." A detailed overview of the components of the proposed MME are presented in Table 1. The weighted categories in the MME provide a summative score on a four-point (0-3) linear scale. The four-point scale provides a necessary level of specificity in the evaluation while maintaining clarity and ease of use for the reviewer (Clemens, Pfitzer, Simmons, Dwyer, Frost, & Olson, 2005). Performance level descriptions are anchored to uniquely defined expectations and avoid narrow prescription. Level 3 is indicative of exceptional instruction and a strong commitment to improving course design and the practice of teaching. Level 2 is indicative of high quality instruction and a commitment to

improving course design and the practice of teaching. Level 1 is indicative of a developing performance that is consistent with the first two years in a college/university position. Sustained performance at Level 1 is generally not consistent with tenure and promotion. Level 0 is indicative of non-performance and is generally not consistent with reappointment.

**Table 1. Weighted data sources and performance levels for the multi-source method for evaluation of teaching (MME)**

	<i>Unsatisfactory:</i> Non-performance; generally not consistent with reappointment (0)	<i>Developing:</i> Needs improvement to be considered for tenure and/or promotion (1)	<i>Commendable:</i> meets expectations for quality instruction; commitment to improving teaching (2)	<i>Exceptional:</i> Exceeds expectations for teaching and shows strong commitment to improving teaching (3)
Student evaluations compared with benchmark instructors; 40% of evaluation	Median score for all evaluations in review is below 10th percentile	Median score for all evaluations in review is between 30 <sup>th</sup> and 10 <sup>th</sup> percentiles	Median score for all evaluations in review is between 70 <sup>th</sup> and 30 <sup>th</sup> percentiles	Median score for all evaluations in review is between 90 <sup>th</sup> and 70 <sup>th</sup> percentiles
Summary of and reflection on student evaluations; 10% of evaluation	No summary or reflection provided	Summary and reflection limited to student evaluation scores and comments	(1) plus summary and reflection related to course difficulty, make-up and content	(2) plus summary and reflection related to teaching methods and styles to retain or develop
Integrated teaching philosophy; 10% of evaluation	No written teaching philosophy provided	Written philosophy statement without evidence of integrated approach	(1) plus instructional artifacts from two classes of integrated approach and two best practices for teaching	(2) plus evidence of three additional best practices for teaching
Professional development of teaching plan; 10% of evaluation	No written plan or evidence of any professional teaching development activities	No written plan, but evidence of at least two activities related to improving teaching	Reflective development plan and evidence of two planned activities to improve teaching	(2) plus one or more additional planned activities related to improving teaching
Mentoring and teaching-related service; 10% of evaluation	No evidence of mentoring or teaching-related service	Summary of one mentoring or teaching-related service activity	Summary of two mentoring or teaching-related service activities	Summary of four mentoring or teaching-related service activities
Reflection on formative external review; 20% of evaluation	No formative external review during the evaluation period	Formative external review, but no substantive reflection	Formative external review with substantive reflection	N/A

## Literature Support for and Explanation of MME Primary Data Sources

### *Student Evaluations*

Numerous studies support the use of student instructional ratings as an effective method of evaluating teaching (Balam & Shannon, 2010; Pan, Tan, Ragupathi, Booluck, Roop, & Ip, 2009; Hoyt & Pallett, 1999; d'Apollonia & Abrami, 1997; Greenwald, 1997; Marsh & Roche, 1984) so long as most students complete the assessment and have a personal stake in the primary learning objectives. Important limitations are also addressed in the literature (Balam & Shannon, 2010; Hoyt & Pallett, 1999; Greenwald, 1997) when student rating systems are poorly constructed and methods of administration may not be standardized<sup>4</sup>. Some rating systems do not take external variables such as class size, student motivation, and course level/discipline into account as the effect of these factors on the overall evaluation is debated. Faculty who assign higher grades may obtain higher student evaluations; yet the impacts of class size and grade leniency on student ratings have been shown to be negligible (Pan, et al, 2009). Alternatively, courses in which students learn more should correlate with higher student performance and grades (Greenwald, 1997; Greenwald & Gillmore, 1997). Another limitation is the inability of students to accurately judge course content (Balam & Shannon, 2010). Given differing viewpoints on how accurately students are able to rate instruction, the student evaluations should represent no more than half of the overall evaluation (Balam & Shannon, 2010; Pan et al 2009; Hoyt & Pallett, 1999; Greenwald, 1997; d'Apollonia & Abrami, 1997; Costin, Greenwald, & Menges, 1971; Marsh, 1984).

### *Weighting of Student Evaluations*

Students, the key stakeholder group in the learning relationship, are given significant voice at 40% representation of the performance evaluation score (see Table 1). A 40% weighting also provides a degree of protection for faculty members in years when low student evaluations are acceptable if the instructor excels in other areas of the evaluation. For example, a particular group of students may lack the context to accurately judge the evaluation criteria (Balam & Shannon, 2010; Greenwald, 1997; Marsh & Roche, 1997). Small class sizes or low response rates from students may make the summative feedback vulnerable to outliers. Additionally, many instructors are developing pedagogical skills in the first or second year of teaching that were not learned in content-specific graduate programs (Cashin, 1990). However, one would not expect a first or second-year teacher to excel in all other areas. This reality is reflected in the Level 1 performance score and supported by Cashin's (1990) conclusion that development does not necessarily imply deficiency. Ongoing development is critical for effective teaching, so it is important that the

---

<sup>4</sup> There are a number of standardized student rating systems available to colleges and universities that have been validated through time and practice. Examples include the Student Evaluation of Educational Quality (SEEQ) system, Student Instructional Rating System (Michigan State University), Student Instructional Report (SIR II™), and IDEA™ Student Rating System (The IDEA Center, Manhattan, KS). These instruments generally score performance by eliciting specific student impressions across multidimensional items such as excellence of course, excellence of instructor, and progress toward learning (d'Apollonia & Abrami, 1997; Cashin, 1992).

evaluation system allows faculty members freedom to try new things and make mistakes (Clegg, 2003).

### *Reflective Teaching Portfolio*

Reflection on the context of their courses as it relates to feedback in the student evaluations (see Table 1) is the faculty member's opportunity to contextualize the experience from his or her point of view. Faculty reflection is, perhaps, the most critical piece of a formative approach to assessment because it has potential to be the greatest influence on faculty development (Kreber & Cranton, 2000). Unless a concept of teaching and learning is formally stated (see Table 1), reflecting on approaches to course design and teaching methods and styles is difficult (Seldin, Miller, & Seldin, 2010; Saphier, Haley-Speca & Gower, 2008). Materials demonstrating attributes of good teaching can also be used to articulate and contextualize aspects of a teaching philosophy. In work sponsored by the American Association of Higher Education (AAHE) and the Education Commission of the States (ECS), a list of best practices in teaching at the undergraduate level was developed for both face-to-face and online settings (Chickering & Ehrmann, 1996; Chickering & Gamson, 1987). By referencing best practices and reflecting on teaching methods and styles, faculty members are able to demonstrate a deliberate plan for teaching and learning in the classroom. Similarly, a teaching development plan is a sign that the faculty member has analyzed his or her own teaching and identified areas for improvement (see Table 1). Utilizing contextualized information from reflections on student evaluations and self-assessments can help inform a teaching development plan that includes both short and long-term goals.

Often, activities that take place outside of the classroom are not evaluated or recognized formally. Service such as mentoring and out-of-class teaching have been shown to positively impact intellectual achievement, overall educational experience, student retention, choice of major among students, GPA, and graduation honors (Astin, 1993) (see Table 1). Additionally, faculty members who are actively engaged in the scholarship of teaching and learning have the potential to significantly impact learning for large audiences of students and instructors.

### *Weighting the Reflective Teaching Portfolio*

Ideally, teaching portfolios should be developed in a collaborative manner to promote the collegial support necessary for the purpose of teaching improvement and support of personnel decisions (Seldin, 1993). While review committees should retain the right to address cursory reflections or submissions with individual faculty members, there is evidence to support the value of a general process of reflection and planning and its impact on teaching (Seldin, 1993). The artifacts and reflections submitted by instructors related to their own teaching make up 40% of the overall teaching performance score – equal to that of student input. The intent of this process was to promote the formative nature of reflective contextualization.

### *Formative External Review*

The third MME component is an external review of teaching and subsequent reflection. There are many benefits for all stakeholders in higher education when faculty invite an external reviewer to evaluate their teaching. This process may also be called “peer review” of teaching. Van Note

Chism (1999) conveys that the process of peer review demonstrates that the act of teaching is valued and worthy of continuous improvement. Peer review also indicates the process of teaching is worthy of scholarly investigation to advance the sharing of new knowledge about best practices. Peer review also elucidates the complexity of teaching while adding to the array of evaluation data utilized to measure teacher effectiveness (Van Note Chism, 1999).

One concern among faculty members is that the process of external review of teaching can be adversarial (Kell & Annetts, 2009). Marsh and Roche (1997) and Marsh (1984) provided evidence that external or peer reviews do not correlate well with student ratings because some peer reviewers were not systematically trained nor asked to rate specific behaviors. Further, questions regarding the competency or expertise of the peer reviewer and practical concerns such as time commitment and the need to assure a valid and reliable process may limit the appeal of this form of evaluation (Van Note Chism, 1999). While format may vary, it is expected that qualified individuals will conduct a formal external review of teaching. Examples include classroom observation or videotaping, peer mentoring reports, small group instructional diagnosis, and course review or instructional consultation. To enhance the formative nature of the process, reflection on the peer review results is encouraged and may be submitted by the faculty member (see Table 1).

### **Methods for Assessing Faculty Perceptions of the MME**

The subject of this research was sensitive as it was about faculty perceptions of a tool used to annually evaluate their job performance, and the results are used to substantiate tenure, promotion, merit based pay, and termination decisions. After IRB Protocol approval was secured in spring 2013, confidential, semi structured interviews (see Table 2), in which the participant selected one of four possible interviewers, were conducted. The process allowed in-depth data collection (Patton, 1990; Van Maanen, 1979) as well as researcher ability to seek clarification if necessary in each unique interview (Fontana & Frey, 1994). In this study, each interview was seen as an “oral report[s] that describe[d] the context and expressions of respondents within their own reality” (Lyde 1999, 6).

**Table 2. Semi-structured interview questions regarding perceptions of the MME**

---

1. Is the *multi-source method* assessment process/tool serving as an effective developmental mechanism for you?
    - (For clarification if necessary) In other words, is the assessment process/tool informing you regarding your own teaching? Please explain.
  2. Are you supportive or opposed to the multi-source method assessment tool, the process, or any part therein? Why or why not?
  3. Of the multi-source method assessment tool, process, or any part therein, what works or does not?
    - (For clarification if necessary) What parts of the multi-source method of assessment portfolio are useful and why? What parts are not useful and why?
  4. Do you believe there are alternative or more effective solutions to the multi-source method assessment tool, process, or any part therein? If yes, please explain.
-

### *Participants*

Participants were tenured and tenure track faculty members at a large Midwestern university. Participants eligible for the study were known to the researchers, as they are colleagues. At the time data were collected, 18 faculty members were eligible. Thirteen elected to participate, including the three researchers. Participation in the study had no bearing on the performance review of any faculty member.

### *Data Collection and Analysis*

Data collection procedures were crafted to minimize risk to participants and ensure trustworthiness and authenticity (Janesick, 1994). Participants were invited to be interviewed individually by the research team member of their choice. Two department faculty members (one tenured and one tenure track) and one faculty member emeritus from outside the department collected data. Interviews and audio recordings were transcribed by the researcher with whom participants chose to interview. Participants were assigned a unique identifier for recording, transcribing, and analyzing data. Each researcher coded all 13 transcripts by memoing followed by multiple meetings to discuss and refine the data clusters into relevant themes (Miles & Huberman, 1994).

### **Results**

The current version of the MME was implemented in 2010. At the time of data collection, three annual review cycles had been completed, and most faculty had been through at least two. The majority of participants perceived some developmental benefit ( $n=11$ ) from completing the MME portfolio. The majority of faculty perceived the structure and multisource nature of the MME useful for prompting them to think about important aspects of their teaching. However, many faculty expressed that they perceived less of a developmental benefit with subsequent submissions, as some items like the teaching philosophy did not significantly change over time. Additionally, faculty expressed discontent with the timing due of the MME portfolio. Submission occurs at the end of a calendar year (accounting for spring, summer, and fall semesters) between the fall and spring academic semesters.

The majority of faculty identified themselves as “supportive” or “not opposed” ( $n=13$ ) to the MME tool as a whole, the process of creating the MME portfolio, or any specific part of the MME. Faculty view the multisource structure of the MME as a strength and that it is a fair tool for a diverse faculty with varying teaching methods and styles. Faculty noted the MME tool in its current form is enough, meaning it assures teaching quality without the risk of becoming overly prescriptive. However, all faculty contextualized their support of the MME or its individual parts by describing aspects of the system that need to be improved. In other words, faculty answered yes to the question about supporting MME, but then supported the answer with “why not” responses. This is consistent with the result that all faculty perceived most MME parts to “work,” or be developmentally valuable in some way to their teaching.

All respondents perceived a need for one or more MME parts to be improved ( $n=13$ ). For example, many faculty members referenced the redundant nature of several parts of the MME, particularly the teaching philosophy, noting the teaching philosophy does not change significantly from year to year. Another common concern about the MME was about the significant workload to produce all of the writing at the end of the calendar year. While many faculty referenced their

goal of completing reflections and other associated MME writings throughout the year, they noted the academic reality of other work taking priority. As a result, the push to complete the MME predominates at a time of the year when faculty are fatigued and busy with prepping for another semester of courses. Finally, most participants (n=11) said “No” or that they do not know of an alternative solution to the current version of the MME.

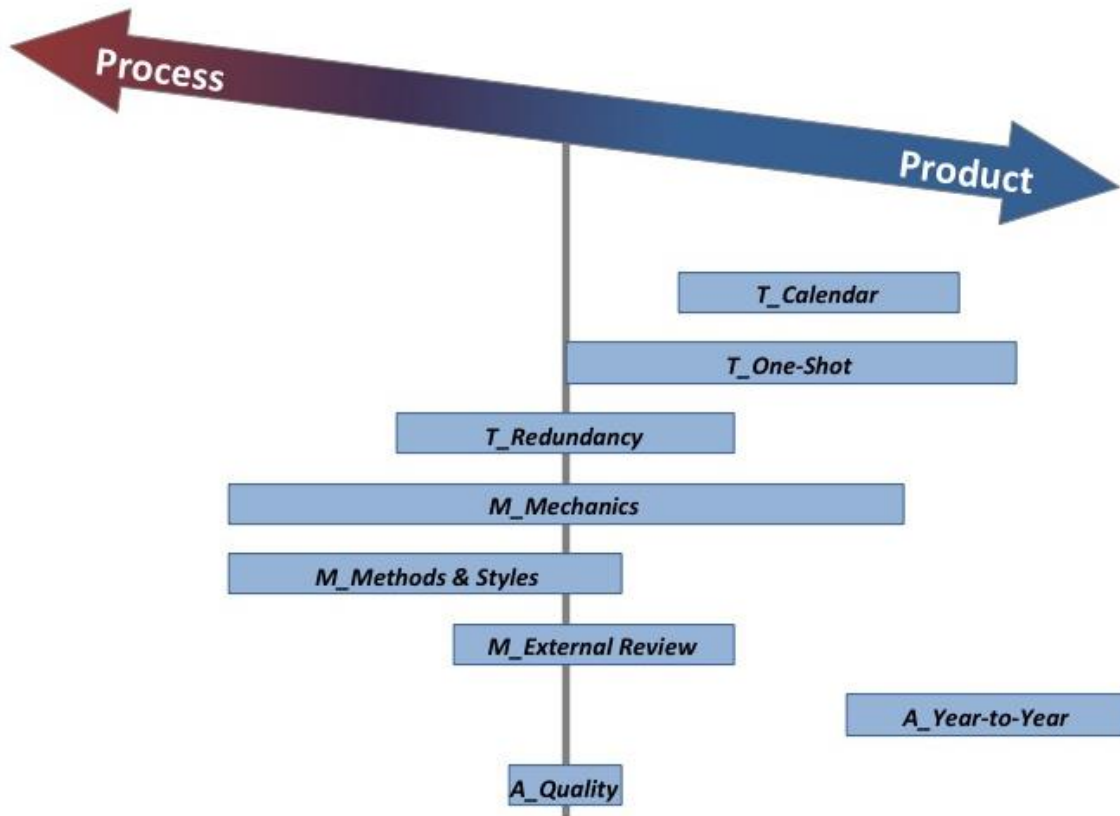
## **Discussion**

Participants recognized and discussed that the MME supports two purposes: 1) develop individual faculty member teaching performance (process/formative); and 2) provide employee performance review information upon which personnel decisions are made (product/summative). However, the primary theme that emerged from the data, consistent with the literature on competing goals for assessment of teaching (Dunn & Mulvenon, 2009; North Carolina Department of Public Instruction, 2008; Van Note Chism, 1999; Raths & Preskill, 1982) is that the MME did not meet its goal of providing a balance between serving as a formative process and creating a product for performance evaluation. The formative process does not necessarily have to include construction of a product to achieve its purpose of teacher development and growth, although it may. Conversely, the development of a product to score performance in teaching is required for personnel decision making, but it does not necessarily ensure a teacher engaged in any formative growth. In the attempt to have the MME meet two distinct purposes, faculty perceive that achieving balance (see Figure 1) is impacted by timing (T) of the assessment, inadequacy of mentoring (M) on the MME, and lack of process for ensuring accountability (A).

### *Timing*

The time of year which the faculty performance review portfolio is submitted shifts the balance of the MME toward a summative product. The reality of time, both in the calendar and academic year, contributes to faculty feeling substantial pressure to produce the portfolio by the deadline. The balance of the MME, as a result, becomes focused on a “one-shot” product instead of a continuous, formative process of reflection that also supports the ongoing development of required, portfolio items such as the Integrated Teaching Philosophy (see Table 1). Another issue associated with timing is redundancy. While some items of the MME may change from year to year, such as the contextualization of student evaluations, other items, such as the teaching philosophy, may remain more consistent from year to year, thereby requiring less reflective effort. Conversely, the teaching philosophy may slowly evolve over longer periods of time and look quite different in the tenth year of teaching compared to the second, and there is no mechanism for accounting for such long term development. For example, the professional development plan generally explains one year of activity. While the MME does provide a summative mechanism for scoring professional development event reporting, its potential to support teaching development over time is limited due to the lack of a requirement for a multiyear or cumulative professional development plan.





**Figure 1. MME use factors that affected the formative and summative balance.**

*Accountability*

Another challenge to the balance between formative process and summative product is accountability, particularly in relation to the professional development plan. The MME provides year to year evidence of a reflective development plan and evidence provided from two activities to improve teaching (as part of the plan). However, there is no requirement to demonstrate that the previous year’s development plan was executed, or, if it was altered, an explanation detailing why. This reality leads to the MME’s balance leaning toward summative product, as there is less focus on long term professional development trajectory and more emphasis on the short term professional development planning and event reporting.

*Mentoring*

A final challenge to the MME balance between a formative, developmental process and the summative product is that of mentoring. First year faculty members are assigned a senior faculty member as a resource and mentor for all aspects of the job. Often, MME mentoring is included, but it is not formalized. The lack of a formal mentoring process for understanding the MME (even for experienced faculty members) shifts the balance toward a summative product and away from a developmental process, as faculty are left to complete the parts of the MME at their own discretion. Given the demands on faculty, and particularly for those learning a new position,

engaging in the work of the MME as a developmental process requires substantial time and reflective effort. First year faculty and faculty who are not trained in pedagogy may be in particular need of mentoring by an experienced faculty member who understands diverse and effective teaching methods and styles. Mentoring by experienced faculty may also ensure understanding of the formative, external review requirements and numerous options for implementation. Mentors may also advise for planning the external review into the academic semester. Finally, mentoring may assist faculty with simple organizational planning and what MME tasks are best to complete at certain times of the year, particularly if developmental benefit is to be realized in the classroom.

## Conclusion and Recommendations

Higher education faculty and administrators may find the results of this study useful, as they reveal practical issues faculty members face using formative processes to improve teaching while simultaneously creating a teaching performance portfolio used for summative evaluation and personnel decision making. The annual performance review process, as in many colleges and universities, is crafted and made policy through a shared governance process among tenured and tenure-line faculty members. Participation in this research study was an opportunity for faculty to provide feedback, in a confidential manner, about how a mutually agreed upon policy for evaluating effective teaching, particularly as a developmental mechanism, was working for them.

While the faculty perceived the MME as a useful tool, they believe it operates primarily to produce a summative product rather than work as a formative process, which counters the goal of the MME policy. The following recommendations should be considered by university faculty and administrators when attempting to increase the formative qualities of a policy or process similar to the MME.

Academic departments should recommend a schedule of due dates that keeps work evenly distributed throughout the year and encourages an ongoing reflection and development cycle. This will not only reduce the proportion of reflective work that occurs when the annual performance review portfolio is due, it will support faculty to reflect (and produce related reflections) during the teaching semesters thereby providing opportunities for faculty to identify challenges and adjust accordingly.

Second, criteria could be added to performance score levels that support faculty demonstration of connectedness among elements or parts of the MME portfolio. For example, how are the student evaluation scores related to or reflective of the teaching philosophy? Or, how does the professional development plan demonstrate a connection to the student evaluation results or the teaching philosophy? Currently, the MME policy only considers reflections related to student and external peer reviews.

While developing additional performance criteria to support demonstrated linkages among the parts of the MME evolve it to provide a more formative process, accountability is also increased from year to year, particularly if an explanation about how the previous year's professional development plan has been addressed or is reflected in a part of the MME such as the student evaluation scores. It is important to note, however, that an effort must be maintained to guard against the overreach into academic freedom by creating an overtly prescriptive process. Efforts must be made to sustain a process that allows faculty to describe how they achieve teaching goals.

Finally, systemic, peer mentoring or *guidance* (not requirement) is needed in the MME policy *and* in the academic department culture. Practices need to be recommended for MME

orientation for new faculty, including examples of how and when to engage in the reflective processes necessary to produce the requisite MME portfolio products. College and university administration and senior faculty should assume mentoring responsibilities to promote a culture where faculty freely and regularly engage in conversations and activities that promote and support teaching development, particularly as it relates to the performance evaluation criteria for products such as the MME portfolio.

### References

- Arreola, R. A. (2000). *Developing a comprehensive faculty evaluation system* (2nd ed.). Bolton, MA: Anker Publishing Company, Inc.
- Astin, A. W. (1993). *What matters in college: Four critical years revisited*. San Francisco: Jossey-Bass.
- Balam, E. M. & Shannon, D. M. (2010). Student ratings of college teaching: A comparison of faculty and their students. *Assessment & Evaluation in Higher Education*, 35, 209-221.
- Cashin, W. E. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology*, 84, 563-572.
- Cashin, W. E. (1990). *Student ratings of teaching: Recommendations for use* (Rep. No. IDEA Paper No. 22). Kansas State University: Center for Faculty Evaluation & Development.
- Chickering, A. W. & Ehrmann, S. C. (1996). *Implementing the seven principles: Technology as lever* (Rep. No. AAHE Bulletin). American Association for Higher Education.
- Chickering, A. W. & Gamson, Z. F. (1987). *Seven principles of good practice in undergraduate education* (Rep. No. 39). American Association for Higher Education Bulletin.
- Clegg, S. (2003). Problematizing ourselves: Continuing professional development in higher education. *International Journal for Academic Development*, 8, 37-50.
- Clemens, P. L., Pfitzer, T. F., Simmons, R. J., Dwyer, S., Frost, J., & Olson, E. (2005). The RAC matrix: A universal tool or a toolkit? *Journal of System Safety*, 41, 14-19.
- Clinton, R. J. (1930). Qualities college students desire in college instructors. *School and Society*, 32, 702.
- Costin, F., Greenwald, A. G., & Menges, R. J. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research*, 41, 511-535.
- d'Apollonia, S. & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52, 1198-1208.
- Dunn, K. E. & Mulvenon, S. W. (2009). A critical review of research on formative assessment:

The limited scientific evidence of the impact of formative assessment in education.  
*Practical Assessment, Research & Evaluation, 14*, 1-11.

Fontana, A. & Frey, J. H. (1994). Interviewing: The art of science. In N. Denzin and Y. Lincoln (Eds.), *Handbook of Qualitative Research* (361-376). Newbury Park, CA: Sage Publications.

Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction.  
*American Psychologist, 52*, 1182-1186.

Greenwald, A. G. & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist, 52*, 1209-1217.

Gustad, J. W. (1961). Policies and practices in faculty evaluation. *Educational Record, 42*, 194-211.

Hoyt, D. P. & Pallett, W. H. (1999). *Appraising teaching effectiveness: Beyond student ratings* (Rep. No. IDEA Paper No. 36). Manhattan, KS: The IDEA Center.

Janesick, V. J. (1994). The dance of qualitative research design: Metaphor, methodology, and meaning. In N. Denzin and Y. Lincoln (Eds.), *Handbook of Qualitative Research*, (209- 219). Newbury Park, CA: Sage Publications.

Kell, C. & Annetts, S. (2009). Peer review of teaching embedded practice or policy-holding complacency? *Innovations in Education & Teaching International, 46*, 61-70.

Kreber, C. & Cranton, P. A. (2000). Exploring the scholarship of teaching. *The Journal of Higher Education, 71*, 476-495.

Lyde, A. (1999). *A qualitative analysis of college females in abusive dating relationships* (unpublished master's thesis). Southern Illinois University Carbondale.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*, 707-757.

Marsh, H. W. & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility." *American Psychologist, 52*: 1187-1197.

McCarthy, M. A., Niederjohn, D. M., & Bosack, T. N. (2011). Embedded assessment: A measure of student learning and teaching effectiveness. *Teaching of Psychology, 38*, 78-82.

McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist, 52*, 1218-1225.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. Newbury Park, CA: Sage

Publications.

North Carolina Department of Public Instruction 2008. *Attributes of Effective Formative Assessment* Washington, D.C.: Council of Chief State School Officers.

Pan, D., Tan, G. S. H., Ragupathi, K., Booluck, K., Roop, R., & Ip, Y. K. (2009). Profiling teacher/teaching using descriptors derived from qualitative feedback: Formative and summative applications. *Research in Higher Education, 50*, 73-100.

Patton, M. Q. 1990. *Qualitative evaluation and research methods*. Newbury Park, CA: Sage Publications.

Raths, J. & Preskill, H. (1982). Research synthesis on summative evaluation of teaching. *Educational Leadership, 39*, 310-313.

Saphier, J., Haley-Speca, M. A., & Gower, R. (2008). *The skillful teacher: Building your teaching skills* (6th ed.). Acton, MA: Research for Better Teaching, Inc.

Seldin, P. (1993). *Successful use of teaching portfolios*. Boston, MA: Anker Publishing Company, Inc.

Seldin, P., Miller, J., & Seldin, C. A. (2010). *The teaching portfolio: A practical guide to improved performance and promotion/tenure decisions* (4th ed.). San Francisco, CA: Jossey-Bass.

Sproule, R. (2000). Student evaluation of teaching: A methodological critique of conventional practices. *Education Policy Analysis Archives, 8*, 1-23

Van Maanen, J. (1979). *Qualitative methodology*. Beverly Hills, Ca: Sage Publications.

Van Note Chism, N. (1999). *Peer review of teaching: A sourcebook*. Boston, MA: Anker Publishing Company, Inc.

Young, S. & Shaw, D. G. (1999). Profiles of effective college and university teacher. *The Journal of Higher Education, 70*, 670-686.