Beyond Student Ratings: "A Whole New World, a New Fantastic Point of View"

Ronald A. Berk, The Johns Hopkins University

You know that professor: the one the students rave about, who always has a huddle of students surrounding her after class, asking questions and chatting, or a line of students outside her door extending along The Great Wall waiting in hope of simply talking to her? The students worship the tile this professor walks on. How do you measure her teaching performance as well as that of all other professors'?

Unfortunately, student ratings have dominated as the primary and, frequently, only measure of teaching performance at colleges and universities for the past four decades (Seldin, 2006). In fact, the evaluation of teaching has been in a metaphorical cul-de-sac with student ratings as the universal barometer. Only recently has there been a trend toward augmenting those ratings with other data sources to broaden and deepen the evidence base (Arreola, 2007; Berk, 2006b; Braskamp & Ory, 1994; Centra, 1993; Knapper & Cranton, 2001; Seldin, 2006).

A Whole New World
One model is a time-tested, industry standard: the 360° multisource feedback (MSF) model, which was developed in management more than half a century ago. Since then, it has gained wide acceptance and over 90% of Fortune 1000 companies use it for formative feedback and summative appraisal decisions (Boyd, 2005).

It works like this. An employee's job behaviors and outcomes are rated anonymously by persons who are most knowledgeable about his or her work (those hierarchically above, below, and on the same level as the employee) to furnish different perspectives (Edwards & Ewen, 1996). This approach taps their collective wisdom to provide a more balanced, complete, and accurate assessment than the traditional single-source, top-down, supervisor-only method. The ratings are compared to self-ratings to give precise feedback to the employee so he or she can plan specific improvements in his or her job performance.

The research on and experience with this 360˚ MSF approach, first used in management and then in healthcare, can be adapted for use in the academy (Berk, 2006, in press). The approach can be easily used for formative and summative decisions about teaching performance and can serve as an appropriate evaluation model for accreditation.

360˚ MSF on Teaching Performance

Critical reviews of strategies to evaluate teaching behaviors suggest a variety of possible raters, including students, self, peers, outside experts, mentors, alumni, employers, and administrators. Furthermore, we can identify 14 potential sources of evidence (Berk, 2005, 2006a, 2006b): (1) student ratings, (2) peer ratings, (3) external expert ratings, (4) self-ratings, (5) videos, (6) student interviews, (7) alumni ratings, (8) employer ratings, (9) mentor's advice, (10) administrator ratings, (11) teaching scholarship, (12) teaching awards, (13) learning outcome measures, and (14) teaching portfolio.

Unfortunately, there is no objective measure of teaching performance; all sources of evidence are fallible. Almost all quantitative and qualitative sources are derived from the 'informed' judgments of students and those persons with whom a professor works. The 360˚ model hinges on the specific decisions about teaching behaviors. The professor is the hub of the ratings. The raters/sources may vary for each decision and change over time as new sources or better instruments are added. For illustrative purposes, a suggested combination of raters and sources will be presented here.

Formative Decisions

Among the 14 sources identified above, which ones would you select to improve your teaching? Which ones provide the most accurate information to pinpoint your strengths, weaknesses, and suggestions on how to improve?

Five of the best sources you could use are: mentor (a level above), peer ratings and video with self/peer feedback (at the same level), student ratings and student interviews (at a level below), plus self-ratings. Different rating scales would be given to the mentor, peer, and students. The professor under review would also complete each of those rating scales. Discrepancies between his or her ratings (self) and those of the other three raters can yield a profile of strengths and weaknesses to pinpoint specific classroom behaviors needing attention. The 360˚ MSF model with these six sources of evidence is shown in Fig. 1 (360° MSF assessment of a professor (formative decisions about teaching):
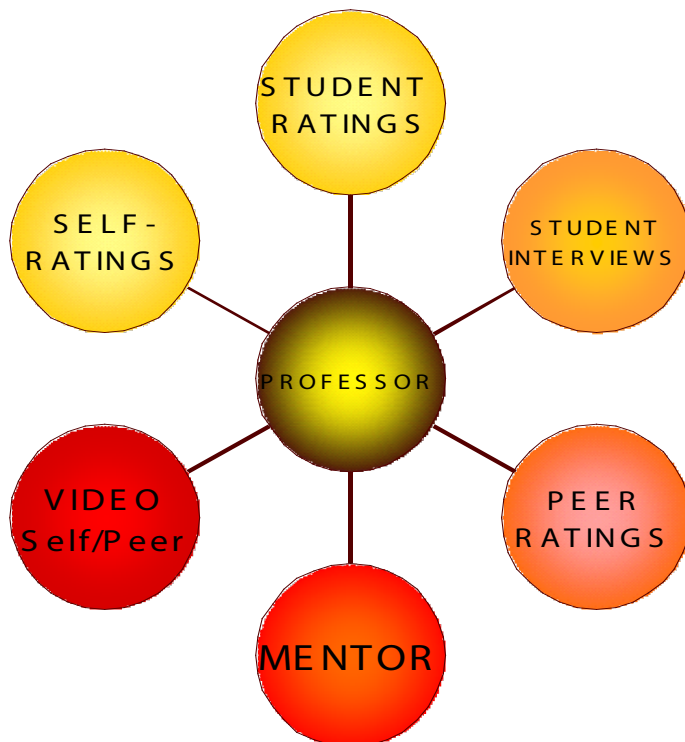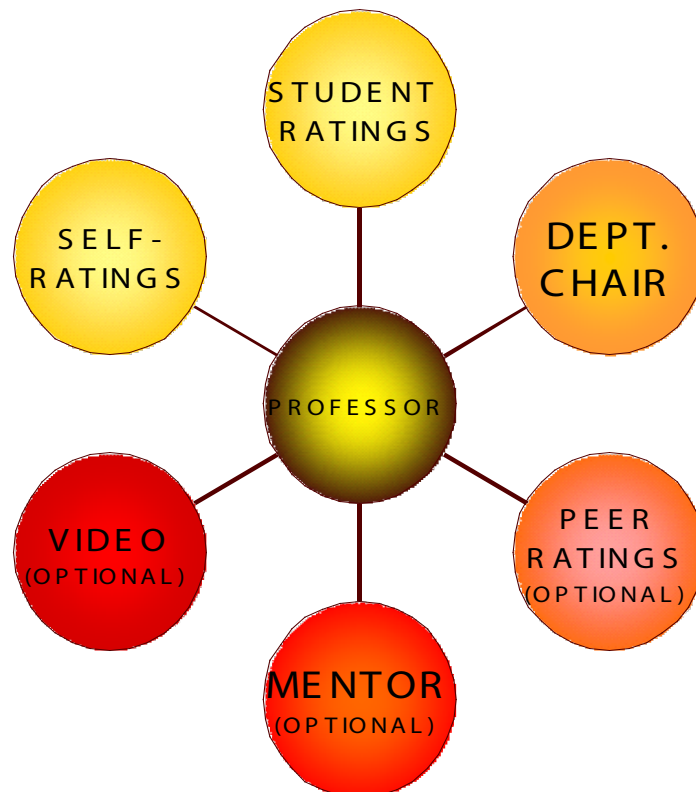
The characteristics of this model are as follows:
1. Professor selects raters and sources of evidence for each;
2. Raters are familiar with professor's teaching behaviors;
3. Those behaviors may be different for each source;
4. Sample group of raters may be large, including students, one or more peers, and a mentor;
5. Different ratings scales with appropriate response options are developed;
6. Quality of many homegrown scales varies from very good to poor but commercial student rating scales are better;
7. Administration of the student rating scales are online or on paper;
8. Data from the different sources are collected at different times during the semester;
9. Feedback from mentor/ peers and from student interviews is immediate (or within days, and from student ratings it can follow in less than two weeks;
10. Professor tracks changes in teaching performance across semesters.

Summative Decisions
Drawing on the 14 sources discussed earlier, which ones would you pick for your department chair or associate dean to determine your annual merit pay or for contract renewal? Which sources provide accurate information on teaching performance, and also collectively converge on a decision that is fair and equitable? Remember: Your teaching career is on the line.

Interestingly, the "best" sources in this instance are almost identical to the ones chosen for formative decisions. They include department chair and mentor ratings (above), peer ratings (optional) and video (optional) (same level), student ratings (below), plus self-ratings. The use of peer and mentor ratings and video feedback for summative decisions should be determined at the discretion of the faculty member; otherwise, it could be a breach of confidentiality (Berk, Naumann, & Appling 2004). The 360°MSF model for this type of summative decision-making is shown in Figure 2:

The salient characteristics of this model are as follows:

1. Faculty and administrator(s) determine the raters and sources of evidence;
2. Multiple raters are chosen for their expertise and to minimize rating bias (Berk, 2006b);
3. Teaching behaviors for each source may differ;
4. Sample group of raters includes the students, one or more peers, mentor, professor (self), and administrator;
5. Separate peer observation reports, different in content and scope from the one shared with the professor, may be submitted to the department chair (Berk et al., 2004);
6. Feedback to the professor by department chair occurs face to face at the end of the year;
7. Department chair tracks changes in teaching performance longitudinally.

There are also summative decisions about promotion and tenure that are often handled very differently than what I am advising in this essay. The department chair may recommend a faculty member for promotion, but the actual analysis of the multiple sources of evidence rests with a committee. The primary source for evaluation would be a teaching portfolio with a variety of elements, including a description of teaching responsibilities, reflective analysis, and artifacts of teaching. These artifacts contain evidence from the aforementioned 14 sources plus examples of course materials and students' work.

Conclusions

Although much has been learned over the 60-year history of faculty evaluation and the 50-year history of the 360˚ MSF model in management, a lot of work is still necessary to realize the true meaning of "best practices" in measuring teaching performance. The model described in this essay is one strategy institutions can use to improve their procedures for rendering fair and equitable decisions.

References

Appling, S. E., Naumann, P. L., & Berk, R. A. (2001). Using a faculty evaluation triad to achieve evidence-based teaching. Nursing and Health Care Perspectives, 22, 247–251.

Arreola, R. A. (2007). Developing a comprehensive faculty evaluation system (3nd ed.). Bolton, MA: Anker.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. International Journal of Teaching and Learning in Higher Education, 17(1), 48-62. (http:/www.isetl.org/ijtlthe.)

Berk, R. A. (2006a, Sept.) A new spin on 360° multisource assessment for teaching/clinical effectiveness. Plenary address presented at the annual meeting of the Association for Medical Education in Europe, Cotone Congessi, Genoa, Italy.

Berk, R. A. (2006b). Thirteen strategies to measure college teaching. Sterling,VA: Stylus.

Berk, R. A. (in review). Using the 360° multisource assessment model to evaluate teaching and professionalism. Medical Teacher,

Berk, R. A., Naumann, P. L., & Appling. S. E. (2004). Beyond student ratings: Peer observation of classroom and clinical teaching. International Journal of Nursing Education Scholarship, 1(1), 1–26.

Boyd, N. M. (2005). 360-Degree performance appraisal systems. In J. Rabin (Ed.), Encyclopedia of public administration and public policy. Oxford, England: Taylor & Francis.

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

Centra, J. A. (1993). Reflective faculty evaluation. San Francisco: Jossey-Bass.

Edwards, M. R., & Ewen, A. J. (1996). 360° Feedback. NY: American Management Association (AMACOM).

Knapper, C., & Cranton, P. (Eds.). (2001). Fresh approaches to the evaluation of teaching (New Directions for Teaching and Learning, No. 88). San Francisco: Jossey-Bass.

Seldin, P. (Ed.). (2006). Evaluating faculty performance. Bolton, MA: Anker.

Ronald A. Berk (Ph.D., University of Maryland, College Park) is Professor Emeritus, Biostatistics and Measurement, The Johns Hopkins University. (www.ronberk.com)

_____