

Faculty Evaluations by Medicine Residents Using Known versus Anonymous Systems

Dayton W. Daberkow II, MD; Charles Hilton, MD, Charles V. Sanders, MD,
Sheila W. Chauvin, MEd PhD

Louisiana State University Health Sciences Center
School of Medicine – New Orleans

Abstract: Purpose. This study examined the extent to which faculty evaluation results differed, based on whether residents were required to submit ratings anonymously or not.

Method. We used a retrospective analysis of existing records representing Internal Medicine residents' evaluation of 51 faculty members in an anonymous and known (non-anonymous) rater system on an inpatient medicine service.

Results. Mean scores for 48 of 51 individuals were lower for anonymous than non-anonymous evaluations. The mean scores were as follows: Anonymous = 5.4 (95 % CI, 5.2-5.6); Non-anonymous = 6.1 (95 % CI, 5.9-6.3). Regression analysis of mean scores for non-anonymous evaluations against those for anonymous evaluations revealed a significant relationship ($r = 0.83$, $p < 0.001$).

Conclusions. Faculty evaluations completed anonymously by residents are significantly lower than those for which resident identities were known. Given the strong, significant relationship between individual faculty members' evaluation ratings from both systems, other factors influencing evaluations should be considered. No subgroups suffered more under anonymous rankings.

Evaluation of faculty by residents is important for ongoing educational program improvement and to maintain program accreditation.^{1,2} For example, Section III, E. 3. of the ACGME institutional requirements states that *the Sponsoring Institution must ensure that residents submit to the program director or to the DIO at least annually confidential written evaluations of the faculty and of the educational experiences.* In addition, career advancement and faculty compensation (e.g., salary and merit pay) are increasingly tied to learners' evaluations of their effectiveness, especially in the contemporary context of accountability, such as where Mission-Based Budgeting models are being used.³ Residents' evaluations of faculty may be collected using systems in which the evaluator is either anonymous or known (non-anonymous).

Faculty members vary in their views and beliefs about whether their evaluations should be obtained using known or anonymous evaluators. For example, some faculty members believe that anonymous evaluations can be influenced by unfair bias and argue that residents should be required to sign their evaluations of faculty members to promote accountability and accurate assessments. Others believe that residents may not be honest in their evaluations of faculty for fear of retribution if they provide criticism or negative feedback. Consequently, their views are

that known evaluators produce inflated evaluation ratings (sometimes referred to in the literature as a "generosity factor" or "halo effect").^{4,5,6} Prior to June 1996, the Internal Medicine residency program at Louisiana State University School of Medicine-New Orleans used a faculty evaluation system in which residents signed their evaluations of faculty teaching effectiveness. Since July 1996, residents' evaluations of faculty have been completed anonymously. Therefore, this study targeted retrospective analysis of faculty evaluation results in both systems to examine the extent to which faculty ratings differed, based on whether residents' evaluations of faculty members were signed or not.

Review of Relevant Literature

Given the conflicting faculty perspectives, we searched the relevant literature to identify and examine empirical evidence of the effects of anonymous and known evaluators on faculty evaluation ratings. Searching the Medline database and using multiple combinations of the following search terms: *residents, evaluation, faculty evaluation, teaching effectiveness*, only one published report was identified that examined the influence of anonymity on faculty evaluation results in medical education. Pelsang and Smith⁷ compared three different methods for gathering residents' evaluations of faculty members' teach-

ing effectiveness. Higher evaluation ratings were observed for individual anonymous questionnaires than for two other methods: 1) large group discussion led by the chief resident with group consensus rating and no faculty present and 2) a representative group discussion with group consensus rating and during which the resident director was present. Numerical ratings differed across the three data collection methods. The large group discussion yielded the lowest ratings, with the representative group in the middle and the anonymous questionnaire producing the highest ratings. Residents were also asked to rank order faculty members, based on perceived teaching effectiveness and the observed rankings varied widely across the three data collection methods.

A related search of the literature produced one report, revealing that most U. S. medical schools use an anonymous system for medical students' evaluation of faculty members' effectiveness as educators, but this was published more than 20 years ago.⁸ We also conducted an additional search of the literature pertaining to learners' assessments of teachers' effectiveness in K-12 and general postsecondary settings using the Education Resources Information Center (ERIC) database (<http://www.eric.ed.gov>). Using the combination of search terms, *faculty* and *teaching effectiveness* and *student ratings*, 356 publications were identified. Of these, several publications were particularly relevant for the study reported here. For example, four studies using control groups were identified and revealed statistically significant differences between faculty evaluation ratings, based on whether anonymous or known raters were used.⁹⁻¹² In these studies, faculty ratings were higher when students' identities were known than when evaluations were submitted anonymously.

Researchers who are well-known for their ongoing work and contributions to studying student evaluation of teaching in higher education have reported consistent findings across several decades that agree on three important points: 1) evaluations in which student identities are known result in higher faculty ratings than anonymous forms, 2) student ratings are higher when forms are completed in the presence of the faculty member being rated, and 3) student ratings are higher when the directions indicate that results will be used for personnel decisions, as opposed to use solely by the faculty member for instructional improvement.¹³⁻¹⁵ Cashin^{14,15} cautions that findings represent common patterns across learners, settings, and institutions and that unique circumstances at a local level should be examined to understand the local influence of any particular variables of

interest. . The supporting evidence and the four studies mentioned previously were based in traditional and general higher education settings, rather than a medical school. Consequently, we cannot assume that the findings from these studies generalize to the residency environment where teaching and learning is predominantly embedded within a work environment (e.g., residents as employees and as learners within the clinical work setting). In addition, teaching/learning occurs in small groups/work teams, usually for very short periods of time (e.g., two- or four-week rotations). As a result, the identity of the resident (evaluator) may be obvious. Residents may avoid submitting low ratings or negative feedback, fearing retribution from faculty who could affect their future, and might consciously or unconsciously inflate evaluations in a known rater system. While written directions do not state that results of faculty evaluations are used for personnel decisions, it is commonly recognized by residents that their collective ratings are considered as part of faculty members' regular performance evaluations, salary and merit pay decisions, and determinations for academic career advancement. For these reasons, we tested the hypothesis that given faculty evaluations available from both anonymous and known evaluation systems, the results of residents' evaluation of faculty teaching effectiveness would be lower when residents submitted ratings anonymously than when their identities on evaluations were known.

Methods

A retrospective, quantitative analysis of residents' evaluations of faculty reflected the research design for this study. First, all known/signed evaluations that were completed July of 1992 to June of 1996 and all anonymous evaluations for July 1996 to June 2000 comprised the data set available for inclusion in the study. Second, we identified faculty members who were included for both the known and anonymous systems for the specified period of time. Third, we set a rule that analyses would include only faculty evaluation data for those individuals who were included in both systems and for whom four or more completed evaluations were available in both systems.

Fifty-one (51) faculty members were identified for which these rules were satisfied. Evaluation data for eleven other faculty members were not included because data were available for three or fewer evaluations in one or both systems. Residents in the categorical, preliminary, and combined internal medicine programs evaluated faculty who had attended on an inpatient medicine service for a one-month rotation at

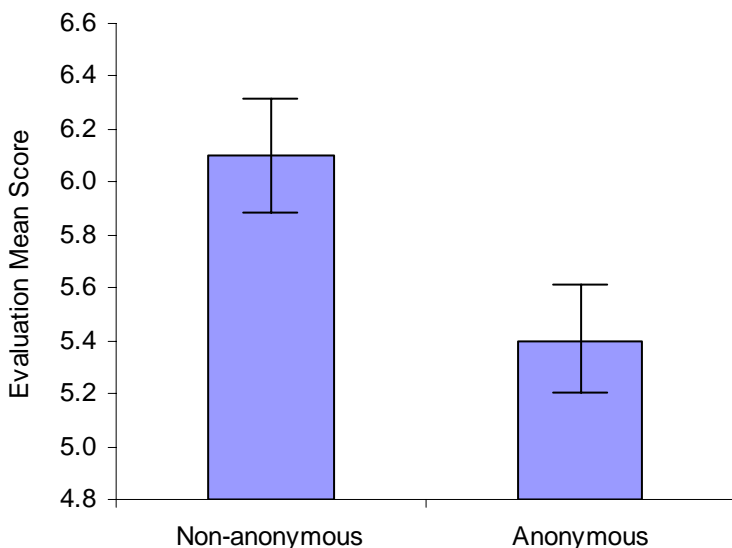


Figure 1. The mean of the known/non-anonymous evaluations compared to the anonymous evaluations of the same faculty, with 95% confidence interval bars. Data are for 51 faculty members from 1992 to 2000.

one our major teaching hospitals. The known evaluations (July 1992-June 1996) were completed monthly and the anonymous evaluations (July 1996-June 2000) for each rotation were completed quarterly. Each one-page form asked residents to score their faculty on overall teaching effectiveness. A seven point rating scale was used in both systems, with 1 = poor clinical teacher and 7 = excellent clinical teacher. We compared the mean evaluation score for overall teaching effectiveness in both systems. Data were expressed as mean values and 95 % confidence intervals were calculated. Linear regression was used

to examine the relationship between faculty members' ratings in known and anonymous evaluation systems.

Results

As shown in Figure 1, the mean number of anonymous and non-anonymous evaluations of each faculty member was 21 (95 % CI, 18-24) and 20 (95 % CI, 17-23) respectively. The mean evaluation score was 6.1 (95 % CI, 5.9-6.3) for known evaluations and 5.4 (95% CI, 5.2-5.6) for anonymous evaluations (maximum possible score was 7 for both groups). The difference in the two groups was significant ($p < 0.001$). Of the 51 faculty members, the mean evaluation score for 48 individuals was lower for anonymous evaluations than for known evaluations. Examination of the descriptive statistics

revealed that faculty evaluation ratings were lower in the anonymous systems than the non-anonymous systems for all individuals, except three. Two individuals had higher mean scores for anonymous raters than for known raters and one person had the same mean score in both systems. Figure 2 shows the results of regression analysis for ratings in the anonymous system on ratings for the known system. Results reflect a strong, positive, and statistically significant predictive relationship ($r = 0.83$, $p < 0.001$).

Table 1
The Mean % Change from Known/Non-anonymous to Anonymous Scores by Sex and Subgroups

Subgroups	Mean % Change	No. of Faculty
All Faculty	14 (11-17)*	51
Male	14 (10-18)	40
Female	15 (7-23)	11
Academic Rank		
Assistant Professor	15 (9-24)	23
Associate Professor	16 (6-16)	10
Professor	11 (8-14)	18
Quartile 1 (Most Effective Teachers)	10 (8-12)	13
Quartile 2	9 (6-12)	10
Quartile 3	10 (6-14)	15
Quartile 4 (Least Effective Teachers)	16 (9-23)	13

* 95% Confidence Interval

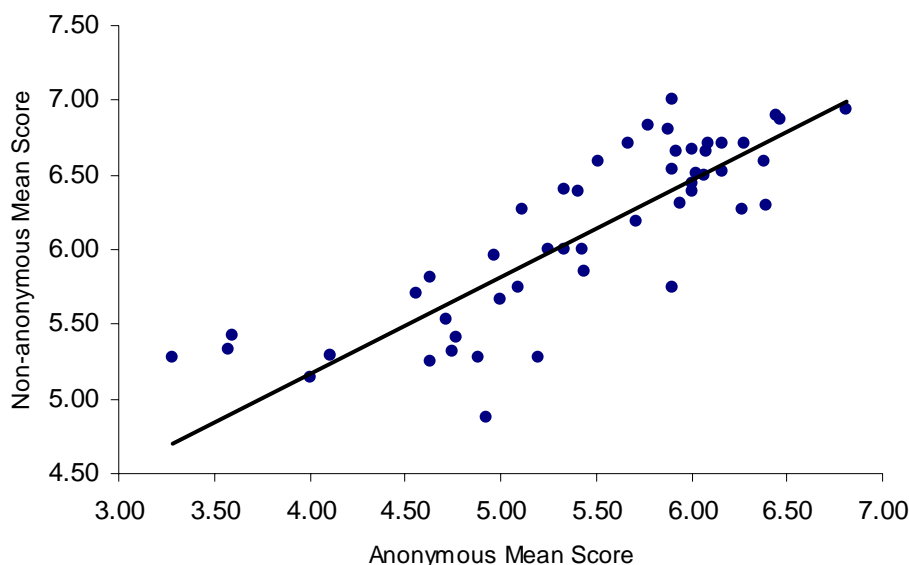


Figure 2. Regression analysis of the mean of residents' known/non-anonymous evaluation scores of faculty members against the mean of anonymous faculty evaluation scores. A significant correlation between the mean of known/non-anonymous scores for each faculty member and the corresponding anonymous mean scores was calculated ($r = 0.83$, $p < .001$).

Supplemental analysis examined the data by sub-groups for gender and academic rank. Analyses were also used to examine the extent to which differences were observed based on the actual ratings, sorting faculty into quartiles based on their mean non-anonymous evaluation scores. The percent change in each faculty's mean evaluation scores for anonymous and known systems was also calculated for each sub-group and the results are summarized in Table 1 and displayed graphically in Figure 3. The mean score for the highest scoring quartile (best clinical teachers) dropped 10% (95% CI, 8%-12%) while the lowest non-anonymous quartile (worst clinical teachers) dropped the most at 16% (95% CI, 9%-23%) and the difference in score change was not statistically significant ($p > .05$). The 2nd and 3rd quartiles dropped 9% (95% CI, 6%-12%) and 10% (95% CI, 6%-14%) respectively. Male scores were 14% (95% CI, 10%-18%) lower in the anonymous system than the known system, and the mean score change for females was similar, 15% (95% CI, 7%-23%). For the faculty rank sub-group analyses, Assistant Professors' mean evaluation scores were 15% (95% CI, 9%-24%) lower in the anonymous system, compared to the known system and results were similar for other faculty ranks: Associate Professors fell 16% (6%-16%), and Professors fell 11% (95% CI, 8%-14%). None of the mean score change differences (known

minus anonymous) in these sub-group analyses were statistically significant.

Discussion

The results of this study are consistent with findings reported in the higher education literature that faculty ratings on anonymous evaluations are lower than those obtained through known evaluator/rater systems. Also, based on the results for this study, there does not appear to be any significant impact for particular sub-groups based on gender, faculty rank, and perceived teaching effectiveness. Our analysis of the data has also yield several interesting insights and questions for future inquiry. However, before we discuss our interpretations and implications, there are clear limitations to this study that we wish to set forth.

First, the study was conducted in a single residency program at one institution. Consequently, the conclusions drawn may not generalize to other residency programs, residents in other specialties, or other academic institutions. Second, the results are based on data that were examined retrospectively and not collected prospectively using an experimental research design in which controlled manipulation of rater anonymity was used. Data were obtained from individuals during two different time periods, so

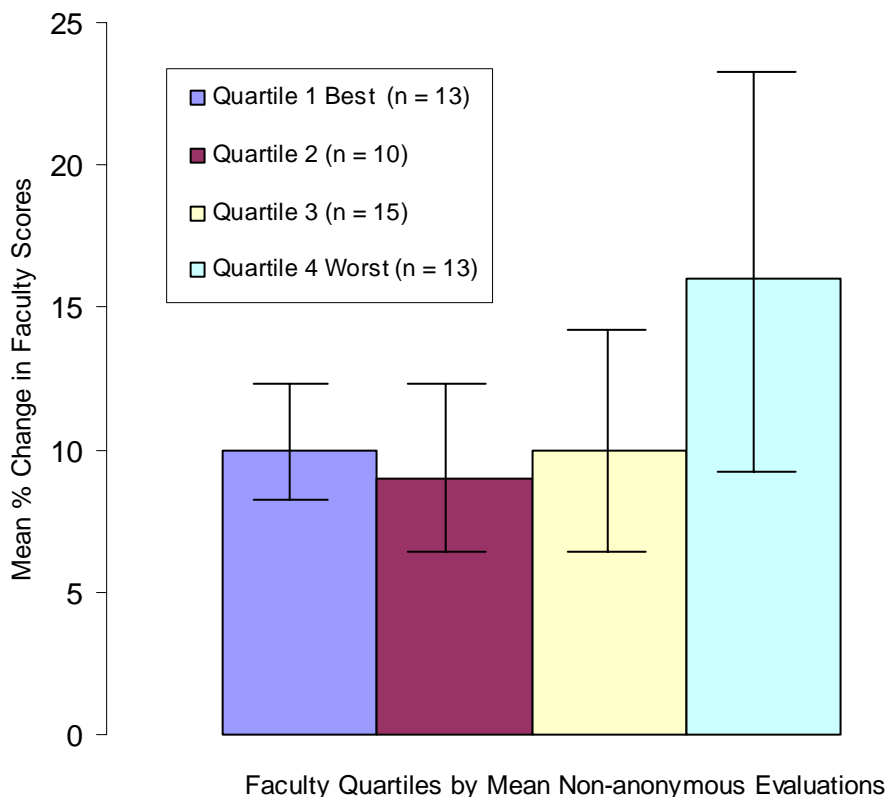


Figure 3. Change in faculty rankings under the known/non-anonymous and anonymous systems. Faculty members were divided in quartiles according to their mean known/non-anonymous scores. The percent change of mean anonymous vs. known/non-anonymous scores was calculated for each faculty member in each quartile. The mean percent change of faculty members by quartile (with 95% confidence interval bars shown) was as follows: Quartile 1 = 10%, Quartile 2 = 9%, Quartile 3 = 10%, Quartile 4 = 16%.

while the faculty members were the same, the residents/raters were different for both periods of data collection. Third, the number of completed evaluations ($n = 4$) for each faculty member evaluated in both systems was small. A larger number would have been desirable. Fourth, the temporal feature of evaluation in each system differed (monthly in the known system and quarterly in the anonymous system). Finally, faculty teaching effectiveness was rated using a global measure, so discrimination ability may have been limited and ratings may have been susceptible to either “halo” (positive bias) or “pitchfork” (negative bias) effects.

Despite the limitations, this study is important because of the increasing significance of faculty evaluation results for program accreditation and faculty career decisions. Few systematic inquiries have been conducted in medical education to examine individual and organizational factors (e.g., known versus anonymous evaluation systems) that may influ-

ence the nature of faculty evaluation ratings. The influence of anonymity on faculty evaluation ratings are consistent with those reported from the larger environment of higher education mentioned previously in this article,¹²⁻¹⁶ but contrast with those reported by Pelsang and Smith.⁷ These authors reported a pattern of increased evaluation ratings corresponding to the order of data collection methods used and suspect that the order and implementation may have contributed to the increasing positive ratings, with the anonymous rating forms being the last method.

In terms of the influence of anonymity on evaluation ratings, Albanese^{5,6} argues that when learner groups are very small, as is often the case for teams on clinical rotations, evaluations submitted anonymously may not be sufficient to avoid concerns about retribution and results may be inflated as well. As was mentioned at the onset of this article, residents may have elevated ratings on faculty evalua-

tions in known systems, because of conscious or unconscious fear of retribution from faculty who could adversely affect the resident's future. For example, residents work and learn with faculty members in small, highly interactive team environments; evaluations are based on short periods of time (e.g., monthly, quarterly), and residents are aware that they may need to call upon these faculty in the future for letters of recommendation to obtain employment when they complete training.

It is interesting to note that while the mean evaluation score differences based on known and anonymous systems were statistically significant in our study, both mean scores were substantially above the midpoint of the seven-point scale (5.4 for anonymous and 6.1 for known systems). This observation might be further explained in terms of another point Albanese makes regarding the discrimination ability of measurements used to evaluation teaching/learning effectiveness. Consequently, while a global measure of teaching effectiveness has been shown to be valid, it is also recognized that it may not be sensitive enough to discriminate adequately and provide diagnostic information to facilitate performance improvement. The results from our study did reflect some differentiation ability of the global measure (i.e., quartiles), but such a measure does not provide any diagnostic benefits in terms of specific areas of strength or needs for improvement. While our study did not focus on these aspects, our analysis and interpretation of findings certainly raise these questions for future inquiry.

We do not know the extent to which these results reflect true teaching effectiveness of faculty members who were evaluated or the extent to which residents in both systems may have been influenced by a "generosity factor" or "halo effect." The generosity factor/halo effect could have been a factor in one or both systems because resident/respondent groups were small (i.e., small teams of residents on any single rotation) or because a global rating of teaching effectiveness was used rather than a multiple item (more diagnostic) measure. If other methods of faculty evaluation for teaching effectiveness had been in place (e.g., peer observation/review of teaching), then we might have been able to triangulate results to address this question. Certainly studies in the future might be designed to incorporate multiple measures and data sources for examining actual teaching effectiveness.

Differences obtained in this study for known and anonymous systems may be due simply to different evaluator groups represented in the two data sets

(1992-1996 and 1996-2000) we used for the retrospective analysis. However, if this were the sole factor contributing to the observed differences, then one would not expect to observe the strong, positive, and statistically significant prediction of anonymous evaluation ratings for observed non-anonymous ratings.

At the onset of our study, we made an assumption that faculty members' teaching effectiveness is demonstrated consistently over time. Specifically, that the targeted faculty members taught at least as well between 1996 and 2000 as they did between 1992 and 1996. As we consider explanations for the observed differences between known and anonymous evaluation systems, it is possible that increasing time and task pressures on clinical faculty may have compromised their teaching performance so that their evaluations would have gone down regardless of the evaluation system used. Another explanation for the drop in evaluation ratings is that faculty members may become complacent in their teaching as they progress through their careers or as they are expected to take on increasing, and sometimes competing, roles. Medical educators recognize commonly the multiple changes and increased expectations for faculty members in recent years.

It is possible that certain aspects of teaching effectiveness have been maintained and improved, while others may have deteriorated. However, because a global measure of faculty teaching effectiveness was used, fine distinctions or examination of specific attributes of teaching effectiveness were not possible in this study. A larger number of evaluation items and specificity of item content to support diagnostic assessment of teaching effectiveness would have facilitated a more detailed explanation regarding the influence of anonymity on faculty evaluation ratings and the exploration of other possible variables that may contribute to resident evaluations of faculty members' teaching effectiveness. Albanese^{5,6} discusses a number of these and provides useful guidance for designing assessment procedures and corresponding studies that might address such concerns.

One other potential explanation warrants further inquiry. That is, there is some evidence that differences in the faculty evaluation ratings obtained for the two time periods (known in 1992-1996 and anonymous in 1996-2000) may be explained, in part, by changes in the amount of time residents spend with faculty members or specific, contextual features of a rotation (e.g., required or elective rotation) that may influence residents' perceptions of faculty teaching effectiveness. Specifically, in a study conducted

in 1987 by Irby, Gillmore, and Ramsey¹⁹ found that the highest ratings for teacher effectiveness were related to faculty members' involvement with residents and students. They also reported that higher teaching effectiveness was perceived in an elective rotation, rather than the required clerkships. Ramsbottom-Lucier and colleagues provided further evidence in their 1994 study that residents' ratings of faculty could be related to the degree of involvement residents perceived of their attending faculty members.²⁰ Consequently, it is likely, and not surprising, that multiple factors influence the ratings assigned by residents to teaching effectiveness on faculty evaluations. The LSU School of Medicine – New Orleans will be implementing a mission-based budgeting model beginning July 2005, so there may be future opportunities to examine yet additional influences (e.g., time and involvement) on faculty members' teaching effectiveness and learners' reported perceptions/evaluations of these performances.

While we cannot conclude that residents' ratings on anonymous evaluations were more accurate assessments of teaching effectiveness than those obtained in the known system, we suspect that mean evaluation scores were lower because faculty members' teaching effectiveness were more honestly rated in the latter system. However, we must also consider that when residents are freed of accountability for their ratings, their personal preferences, feelings, dislikes or other irrelevant factors may yield lower ratings than when their anonymity is not assured. Given the global measure of teaching effectiveness examined in this study, the influence of such factors (positive or negative) could more easily impact the accuracy of ratings. Finally, although the mean evaluation scores were lower in the anonymous system, it is important to note that a negative impact was not observed for any specific subgroup (e.g., gender, faculty rank, performance/score quartile). One may conclude that since adequate separation was already achieved in the known system, then an anonymous system isn't needed unless it has an added benefit for program directors and medicine departments. For example, from a measurement perspective, if the goal of faculty evaluations is solely to distinguish good teachers from bad teachers, then global numerical scores may be sufficient for sorting individuals based on overall teaching effectiveness and for making decisions about teaching/supervision assignments, allocation of resources for specific roles, and quality assurance and accountability (e.g., accreditation requirements). For example, the ACGME requires programs to have formal mechanisms for annual faculty evaluations of teaching staff by residents. All ACGME-accredited programs must use these evalua-

tions for faculty-member counseling and for selecting faculty members for specific teaching assignments. Outcomes from resident evaluations such as the number of comments, feedback to faculty members, and how this impacts a department and its teaching effectiveness are important for all residency programs. Consequently, global measures may be sufficient to satisfy accountability and quality assurance requirements, but they may not be sufficient to facilitate ongoing individual professional development and educational program improvement.

Regardless of the form of evaluation instruments, assurance for validity and reliability appropriate to its purpose is important. A number of studies are available regarding instrument development and associated validity and reliability and there is evidence that global measures can produce valid and reliable results. For example, (e.g., Williams and colleagues at the University of Michigan Medical School) demonstrated that a global measure can be as valid and reliable as a more detailed and discriminating, multiple item measure.²¹

If a goal for evaluating teaching effectiveness is to also enhance the educational program and promote continuous professional development of faculty members in the art and science of teaching, then multiple-item instruments and multiple methods, such as the use of both quantitative (scores) and qualitative (content analysis of narrative comments), seem necessary to provide specific, diagnostic assessments about teaching effectiveness. More rigorous evaluation of teaching effectiveness might also include multiple data sources (e.g., students, residents, disciplinary peers, faculty development specialists) to support triangulation of results and facilitate appropriate interpretation and decision-making. Such rigor and multi-dimensionality seems particularly important in light of situations where such outcomes carry increasing significant for career and compensation decisions, such as in mission-based budgeting contexts and environments where scholarly work in teaching/learning is considered more seriously for tenure and promotion decisions.

Specific assessment results that are observation-based and grounded in specific criteria are necessary to provide feedback effectively in formative faculty professional development and formal faculty evaluation processes. Some residency programs have demonstrated through systematic investigation that appropriate feedback to faculty can be an effective tool for improving teaching performance.^{22, 23} For example, in a study of 44 pediatrics faculty at the Medical College of Wisconsin, residents and students evalua-

ated faculty members' teaching effectiveness.²⁴ Faculty members were then randomly assigned to either a control group or a group receiving specific feedback. Results showed that those receiving feedback had significantly higher increases in evaluation ratings over time than those in the control group (no feedback).

In conclusion, the results of this study seem consistent with results obtained in higher education regarding the influence of anonymity on ratings of faculty teaching effectiveness. Other factors that may influence learner ratings of faculty teaching effectiveness, of which some may be particular to residency education, need to be further examined across specialties, settings, and institutions. Consequently, additional studies are needed to confirm or refute alternative explanations presented in this article and to explore the extent to which they may affect global versus diagnostic forms evaluating teaching effectiveness. The design of a multi-specialty, multi-institutional study could contribute significantly to examining the influence of anonymity and other potential factors on learner evaluations of faculty members' teaching effectiveness. This study has also raised a number of considerations and potential areas for further study regarding the methods of faculty evaluation and how assessment/evaluation results are used for individual and program/organizational development and improvement. Focused and large-scale studies could contribute significantly to identifying the extent to which such evaluation practices add value to the educational processes and outcomes.

As academic medical centers are being held more accountable to the public for quality of care, teaching faculty can expect that their institutions will increase their focus and efforts on faculty members' effectiveness as teachers. With the current expansion of scholarship expectations to include scholarly teaching and learning and the increased use of teaching portfolios/dossiers in annual faculty evaluation and promotion/tenure processes, evaluations of faculty members' teaching is becoming a high stakes enterprise more than ever. Consequently, more studies of the faculty evaluation process and outcomes are needed.

Acknowledgements

The authors wish to acknowledge and thank the following colleagues in the Office of Medical Education Research and Development (OMERAD) at LSUHSC-New Orleans: Tong Yang, MD, MS for his consultation regarding statistical analysis and

Geoff Wiggins, MS for his assistance with conducting systematic searches of the professional literature.

References

1. Accreditation Council for Graduate Medical Education, Institutional Requirements, http://www.acgme.org/acWebsite/irc/irc_IRCp_r703.asp [Accessed April 1, 2005].
2. Graduate Medical Education Directory 2002-2003. 86th ed. Chicago, IL: American Medical Association, 2002.
3. Watson RT, Romrell JJ. Mission-based budgeting: Removing a graveyard. *Acad Med*, 1999, 74: 627-640.
4. Mehrens W, Lehmann I. *Measurement and Evaluation in Education and Psychology*, 4th Ed. NY; Holt, Rinehart and Winston, 1991.
5. Albanese MA. Challenges in user rater judgments in medical education, *J Eval Clin Pract*, 2000; 6: 305-319.
6. Albanese MA. Rating educational quality: factors in the erosion of professional standards. *Acad Med*, 1999; 74: 652-658.
7. Pelsang RE, Smith WL. Comparison of anonymous student ballots with student debriefing for faculty evaluations. *Med Educ*, 2000; 34: 465-467.
8. Smith SR, Paulen LJ. Use of anonymous student evaluations of faculty members in U.S. medical schools. *J Med Edu*, 1984; 59: 196-97.
9. Argulewicz EN, O'Keefe T. An investigation of signed vs. anonymously completed ratings of high school student teachers. *Educ Res. Q*, 1978; 3: 39-44.
10. Cheong, GSC. Students' evaluation of instructors: before and after the examination, names identified versus anonymous. *Can J Higher Educ*, 1979; 9: 80-86.
11. Ronning RR, Walsh UR. Effects of student anonymity-nonanonymity on the factor structure of a teacher rating form. *Res Higher Educ*, 1977; 6: 363-371.

12. Stone EF, Spool MD, Rabinowitz S. Effects of anonymity and retaliatory potential on student evaluations of faculty performance. *Res Higher Educ*, 1977; 6: 313-324.
13. Braskamp LA, Ory, JC. *Assessing Faculty Work: Enhancing Individual and Institutional Performance*. San Francisco: Jossey-Bass, 1994.
14. Cashin WE. *Student Ratings of Teaching: The Research Revisited*, Idea Paper No. 32. Education Resources Information Center, 1995; ED 402 338, <http://www.eric.ed.gov> [Accessed April 1, 2005].
15. Cashin WE. *Student Ratings of Teaching: A Summary of the Research*, Idea Paper No. 20. Education Resources Information Center, ED 302567, <http://www.eric.ed.gov> [Accessed April 1, 2005].
16. Centra JA. *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness*. San Francisco: Jossey-Bass, 1993.
17. Feldman KA. The significance of circumstances for college students' ratings of their teachers and courses. *Res in Higher Ed* 1979; 10: 149-172.
18. Marsh HW, Dunkin M. Students' evaluations of university teaching: A multidimensional perspective. IN JC Smart (Ed.). *Higher Education: Handbook of Theory and Research* (Vol. 8, pp. 143-223). New York: Agathon, 1992.
19. Irby DM, Gillmore GM, Ramsey PG. Factors affecting ratings of clinical teachers by medical students residents. *J Med Educ*, 1987; 62: 1-7.
20. Ramsbottom-Lucier MT, Gillmore GM, Irby DM, Ramsey PG. Evaluation of clinical teaching by general internal medicine faculty in outpatient and inpatient settings. *Acad Med*, 1994; 69: 152-154.
21. Williams BC, Lizelman DK, Babbott SF, Lubitz RM, Hofer TP. Validation of a global measure of faculty's clinical teaching performance. *Acad Med*, 2002; 77: 177-180.
22. Cohan RH, Dunnick NR, Blane CE, Fitzgerald JT. Improvement of faculty performance: efficacy of resident evaluation. *Acad Rad*, 1996; 3: 63-7.
23. Whitman N, Schwenk T. Faculty evaluations as a means of faculty development. *J Fam Prac* 1982; 14: 1097-101.
24. Schum T, Yindra KJ. Relationship between systematic feedback to faculty and ratings of clinical teaching. *Acad Med*, 1996; 71: 1100-2.

Correspondence

Dr. Dayton Daberkow II
LSUHSC Internal Medicine Residency Program
1542 Tulane Ave.
New Orleans LA 70112-2822

Phone: 504-568-6255
Fax: 504-568-7885
Email: ddaber1@lsuhsc.edu