

Clinician-teachers' Self-assessments Versus Learners' Perceptions

Donna M. Windish, MD, Amy M. Knight, MD, Scott M. Wright, MD

Understanding how clinician-teachers' self-assessments compare to learners' impressions can serve to help educators place each of these evaluations in the appropriate context. Past participants of the Johns Hopkins Faculty Development Program and other physician-teachers were surveyed in 2002 regarding their teaching skills and behaviors. We surveyed their learners to compare teacher and learner assessments of teaching proficiency, behaviors, enjoyment, and career satisfaction. In each area, learners' ratings were statistically significantly higher than their teachers' self-ratings. Though it is unclear whether teachers' or learners' assessments are a more accurate reflection of the truth, the more positive learner ratings should promote self-confidence in clinician-educators regarding their teaching abilities.

KEY WORDS: physician-teachers; teacher evaluations; self-assessments.

J GEN INTERN MED 2004;19:554-557.

"It goes without saying that no man can teach successfully who is not at the same time a student."

—Sir William Osler, 1905

A physician's ability to recognize personal strengths and weaknesses is important to their professional development in both clinical and teaching roles. This is particularly relevant after formal training has ended.¹ Accurate self-assessment is central to a physician's professional growth,^{2,3} and may correlate with global self-esteem.^{4,5} Yet the accuracy of self-assessment in medicine has been variable when compared to more objective measures.⁶

Common assessments of a teacher's merit and effectiveness are learners' impressions of the teacher and improvement in skills, attitudes, and behaviors that result from the teacher-learner interaction. While several studies have assessed learner evaluation of clinical teaching,⁷⁻¹² few studies have compared faculty self-assessments with learner estimations.¹³

Received from the Division of General Internal Medicine (DMW), Johns Hopkins University School of Medicine; and Division of General Internal Medicine (AMK, SMW), Johns Hopkins Bayview Medical Center, Baltimore, Md.

This paper was presented at the Society of General Internal Medicine's National Meeting in Vancouver, Canada, April 29 to May 5, 2003, as "A Comparison of Teachers' Self-assessments with Learners' Perceptions."

Address correspondence and requests for reprints to Dr. Windish: Johns Hopkins University School of Medicine, 1830 East Monument Street, Room 8038, Baltimore, MD 21287 (e-mail: dwindish@jhmi.edu).

This study was conducted to better understand how teachers' self-assessments of their skills and abilities compare to learners' perceptions. The results may provide insight into the meaning of teachers' self-assessments and may help in our understanding of the differences between teacher self-assessment and learner opinions.

METHODS

Study Design and Sample Selection

In 2002, 269 past participants in the Johns Hopkins Faculty Development Program and a comparison group of 141 physician-teachers were surveyed regarding their teaching skills and behaviors, as discussed elsewhere (AMK et al., unpublished data, 2003). Of 245 teachers responding by October 2002, 162 indicated that they had taught or mentored during the prior year. These respondents provided 410 names of learners with whom they had worked closely. We surveyed those learners who 1) had legible, complete, and accurate contact information, and 2) were located in the continental United States. If three or more learners were named, two individuals were randomly selected and contacted. This left 217 learners for inclusion in our study. The Johns Hopkins Bayview Institutional Review Board approved the study protocol.

Survey Content and Administration

The survey instrument rated teachers in 4 areas: 1) teaching proficiency, 2) teaching behaviors, 3) teaching enjoyment, and 4) career satisfaction, using 4- and 5-point Likert scales. Information was collected regarding teacher academic rank, learner level of training, gender, duration of teacher-learner relationship, and teaching settings. Learners also rated their teacher's overall teaching ability compared to other teachers they encountered in the previous year. Questionnaires were mailed (172) or e-mailed (45) to learners in October 2002. Follow-up contacts were made to encourage full participation.

Data Analysis

Comparisons between the teachers' and learners' responses were done using the Wilcoxon signed-rank test. Because multiple comparisons were performed, statistical significance was set at $P < .002$ using a Bonferroni correction. Subgroup analyses looking at gender, learner level of training, overall teacher ratings, and previous faculty development training were done using a 2-sample Wilcoxon rank-sum (Mann-Whitney) test. All analyses were performed using Intercooled Stata, version 7.0 (Stata Corporation, College Station, Tex).

Table 1. Characteristics of 117 Clinician Teachers and Their 173 Responding Learners

	<i>n</i> (%)
Teacher Characteristics	
Gender	
Female	76 (65)
Male	41 (35)
Academic Rank	
Instructor	13 (11)
Assistant professor	56 (48)
Associate professor	27 (23)
Professor	5 (4)
Other	16 (14)
Johns Hopkins Faculty Development Trained	
Yes	135 (78)
No	38 (22)
Learner Characteristics	
Gender	
Female	89 (51)
Male	84 (49)
Highest Level of Training While Interacting with the Teacher	
Medical student	33 (19)
Resident	95 (55)
Fellow	17 (10)
Graduate student	7 (4)
Physician completed training	12 (7)
Other	9 (5)
Settings in Which Teacher Interacted with the Learner	
House staff or resident clinic	89 (51)
Morning report or noon conference	86 (50)
Inpatient ward	85 (49)
Teacher's own office/practice	83 (48)
Small group setting (other than above)	76 (44)
Faculty development program or workshop	24 (14)
Mean Duration of the Teacher-Learner Relationship, $y \pm SD$	2.7 \pm 2.1

RESULTS

Of 217 learners surveyed, 173 (80%) returned the questionnaire (Table 1). These learners assessed 117 teachers, 48% of whom were assistant professors. Residents comprised 55% of learner respondents. The teacher-learner relationship averaged 2.7 years (range 0.25 to 20 years). Most respondents had observed teachers in multiple settings (mean 2.9 ± 1.4 settings).

Teachers' Self-assessments Versus Learners' Perceptions

Teachers rated their teaching proficiency lower than their learners' assessments in all teaching skill areas assessed. The largest discrepancies were seen for mentoring (median 3 [interquartile range (IQR) 3 to 4] vs 5 [IQR 4 to 5], where 1 = poor, 5 = excellent), giving feedback (3 [IQR 3 to 4] vs 4 [IQR 4 to 5]), eliciting feedback (3 [IQR 2 to 3] vs 4 [IQR 3 to 5]), and handling conflict (3 [IQR 3 to 4] vs 4 [IQR 4 to 5]; all $P < .001$).

For all teaching-related undertakings (precepting, mentoring, lecturing, developing curricula, and leading

small groups), teachers noted that they derived less enjoyment from these activities than the learners perceived (2 [IQR 2 to 3] vs 3 [IQR 3 to 3], where 0 = not at all, 3 = a lot; all $P < .001$).

Compared to learners, teachers reported that they less frequently performed positive teaching behaviors in all areas assessed (Table 2; most $P < .001$). Finally, learners judged their teachers' career satisfaction to be higher than that which was described by the teachers (5 [IQR 4 to 5] vs 4 [IQR 4 to 5], on a 5-point scale, 1 = strongly disagree, 5 = strongly agree; $P = .001$).

Twenty-two percent of learners rated their teacher as the best teacher they had worked with in the previous year. Forty-eight percent ranked their teacher in the top 3. The remaining 30% placed them in the top half.

Because 56 (48%) teachers had two learners who evaluated them, we randomly eliminated one of the learners and repeated the analyses; this produced no significant differences in any of the results.

Subgroup Analyses

Gender of Learner and Teacher. Female learners felt teachers enjoyed developing curricula and leading small groups more than male learners reported ($P < .02$). They also assessed their teachers' behavior more favorably in effectively helping the group meet goals and allowing learners to figure things out themselves ($P < .05$). Female learners were not significantly more critical of their teachers than male learners for any question.

Learners rated male teachers higher in overall teaching skills and in effectively helping the group meet goals ($P < .01$). Male teachers were also viewed more favorably in overall teaching ability ($P = .03$). They were not appraised significantly lower than female teachers for any domains explored.

Learner Level of Training. When resident physician opinions were compared to those of other learners, they were noted to regard teachers as less proficient in 1) mentoring, 2) role modeling, and 3) giving feedback (all $P < .05$). They also felt teachers less frequently asked learners what they would like to get out of a teaching encounter ($P < .01$). Finally, residents judged their teachers' overall teaching ability less favorably than did other learners ($P = .001$).

Faculty Development Training. Analyses were performed to look for differences in responses among faculty who had and had not participated in the Johns Hopkins Faculty Development Program. There were negligible differences (not statistically significant) with regard to teacher-learner comparisons. However, subtle differences were seen in teacher self-assessments. Teachers without prior faculty development training had higher self-assessments in 4 teaching proficiencies: giving lectures and presentations, mentoring, handling conflict, and leading others (all $P < .02$). Conversely, teachers with faculty development

Table 2. Comparison of Teacher and Learner Assessments Regarding the Frequency with Which the Teacher Performed the Following Behaviors*

Teaching Behavior	Teachers'		Learners'		P Value [§]
	Self-assessment Median (IQR) [†] Percent [‡]		Rating Median (IQR) [†] Percent [‡]		
Effectively helped the small group to meet goals	3 (3 to 3)	83	4 (3 to 4)	96	< .001
Let learners know his/her limitations as a teacher	3 (2 to 3)	59	3 (3 to 4)	83	< .001
Asked learners what they would like to get out of an encounter	3 (2 to 3)	64	4 (3 to 4)	84	< .001
Let learners figure things out themselves, even if it meant struggling	3 (2 to 3)	55	3 (3 to 4)	87	< .001
Expressed concern or support for learners when they were struggling	3 (3 to 4)	83	4 (3 to 4)	94	< .001
Would try to detect the emotional response of the learners to various situations, and discuss this if appropriate	3 (2 to 3)	67	3 (3 to 4)	83	< .001
Spent time building supportive relationships with the learners	3 (3 to 4)	79	4 (3 to 4)	93	< .001
Changed a learning plan based on learner's needs	3 (2 to 3)	68	4 (3 to 4)	86	< .001
Helped learners identify resources that are available to meet their learning needs	3 (3 to 3)	79	4 (3 to 4)	92	< .001
Worked with the learners to establish mutually agreed upon goals, objectives, and ground rules	3 (2 to 3)	66	4 (3 to 4)	88	< .001
In precepting or one-on-one teaching, assessed and focused on the learner's needs rather on his/her own agenda	3 (2 to 3)	70	4 (3 to 4)	94	< .001
In precepting or one-on-one teaching, challenged learners to think about alternative management approaches	3 (2 to 3)	79	4 (3 to 4)	94	< .001
In feedback sessions, focused on specific areas needing improvement, rather than making generalizations	3 (3 to 3)	75	4 (3 to 4)	91	< .001
Starts feedback session by asking learners to assess their performance	3 (2 to 4)	66	3 (3 to 4)	81	.35

* A 5-point Likert scale was used (0 = never, 1 = rarely, 2 = sometimes, 3 = frequently, and 4 = always).

[†] IQR, interquartile range of responses.

[‡] Percent reporting frequently or always performs the teaching behavior.

[§] P values are for the differences in responses between teachers and learners using the Wilcoxon signed-rank test.

training rated themselves higher for select teaching behaviors: asking learners what they would like to get out of an encounter, establishing mutually agreed upon goals, and focusing on the learner's needs (all $P < .01$).

Overall Teacher Rating. Teachers who were judged to be among the 3 best teachers encountered in the previous year were rated statistically more favorably in all but one variable: letting learners know his/her limitations as a teacher ($P = .06$).

DISCUSSION

Prior studies that have evaluated proficiency and quality of physician-teachers have focused primarily on learner opinions. Our study compares teachers' self-assessments with learners' perceptions across a multitude of competencies and characteristics including teaching skills, behaviors, enjoyment, and career satisfaction. In each of these areas, teachers' self-assessments were less favorable than their learners' appraisals.

Several questions arise from the differences observed. Are teachers too self-critical or are learners overly generous in their ratings? Are teachers adequately aware of their strengths and weaknesses? Should teachers be as critical of their abilities if learners are not? Are teachers' lower self-evaluations indicative of depleted self-concept and self-esteem? Might learners be more skilled and perceptive

evaluators because they routinely evaluate teachers, whereas teachers rarely evaluate themselves? These questions, as well as others, can only be addressed by additional empiric work in the area.

Though most subgroup analyses yielded similar results, some differences became apparent. First, gender was noted to have some bearing on the results. Female learners were more positive in their teacher assessments and male teachers were rated more favorably by learners. Though it is unclear why gender should influence assessments, perhaps these trends should be considered when interpreting faculty evaluations. Second, residents judged their teachers' skills and behaviors less favorably than other learners. In doing so, their assessments were a closer approximation to those of their teachers. Ullian et al. observed that as residents progress in their training, their evaluation of teaching changes.¹⁴ While first-year residents perceive their physician-teachers as experts, senior residents consider teachers more as colleagues. Thus, residents' opinions of their teachers, particularly as they mature, may be equivalent to peer evaluation, and as such, might represent a more accurate assessment of a teacher's performance and effectiveness.

In recent years, clinician-educators have expressed anxiety over promotion difficulties.¹⁵ Peer assessment has been recognized by promotion committees as a critical measure in the evaluation process.¹⁶ Though learner evaluations are also considered, less weight and credence are

placed on them.¹⁶ Because our study shows that learners rate clinician-educators higher than these teachers appraise themselves, when being considered for promotion, clinician-educators may wish to advocate that a greater emphasis be placed on learner evaluations, rather than discounting them as being inferior.

Limitations of this study should be considered. First, learners were self-selected by their teachers, which may account for higher learner ratings. However, extensive interactions between the teachers and learners occurred (both in duration of the relationship and in number of teaching settings), which almost certainly resulted in a more informed assessment. Second, physicians' teaching skills were not directly observed by an impartial third party so as to allow comparisons of the evaluations with a gold standard. Third, a small number of learners evaluated each teacher. Some studies suggest that many learners need to evaluate teachers over several years to provide the most accurate ratings.^{17,18} Finally, only a small number of teachers who were not faculty development trained provided learner information. As such, comparisons attempting to explore differences between those who had and who had not participated in faculty development were restricted.

Though it is unclear whether teachers' or learners' assessments are a more accurate reflection of the truth, these results should give clinician-educators greater insight into how self-assessment of their teaching skills compares with their learners' assessments. The results have the potential to improve the self-esteem of clinician-educators and their confidence in their teaching abilities.

Dr. Wright is an Arnold P. Gold Foundation Associate Professor of Medicine. The authors are indebted to Dr. Marie Diener-West for her suggestions.

REFERENCES

1. Wooliscroft JO, TenHaken J, Smith J, Calhoun JG. Medical students' clinical self-assessments: comparisons with external measures of performance and the students' self-assessments of overall performance and effort. *Acad Med.* 1993;68:285-94.
2. Arnold L, Willoughby TL, Calkins EV. Self-evaluation in undergraduate medical education: a longitudinal perspective. *J Med Educ.* 1985;60:21-8.
3. Zabarenko RN, Zabarenko LM. Teaching psychological medicine on hospital rounds: a liaison experiment. *Int J Psychiatry Med.* 1978;8:325-34.
4. Carmel S. The professional self-esteem of physicians scale, structure, properties, and the relationship to work outcomes and life satisfaction. *Psychol Rep.* 1997;80:591-602.
5. Rosenberg M, Schooler C, Schoenbach C, Rosenberg F. Global self-esteem and specific self-esteem: different concepts, different outcomes. *Am Sociol Rev.* 1995;60:141-56.
6. Gordon MJ. Self-assessment programs and their implications for health professions training. *Acad Med.* 1992;67:672-9.
7. Donnelly MB, Wooliscroft JO. Evaluation of clinical instructors by third-year medical students. *Acad Med.* 1989;64:159-64.
8. Irby D, Rakestraw P. Evaluating clinical teaching in medicine. *J Med Educ.* 1981;56:181-6.
9. Williams BC, Pillsbury MS, Stern DT, Grum CM. Comparison of resident and medical student evaluation of faculty teaching. *Eval Health Prof.* 2001;24:53-60.
10. Kendrick SB, Simmons JMP, Richards BF, Roberge LP. Residents' perceptions of their teachers: facilitative behavior and the learning value of rotations. *Med Educ.* 1993;27:55-61.
11. Ramsey PG, Gillmore GM, Irby DM. Evaluating clinical teaching in the medicine clerkship: relationship of instructor experience and training setting to ratings of teaching effectiveness. *J Gen Intern Med.* 1988;3:351-5.
12. Irby DM, Gillmore GM, Ramsey PG. Factors affecting ratings of clinical teachers by medical students and residents. *J Med Educ.* 1987;62:1-7.
13. Hewson MG, Copeland HL. What's the use of faculty development? Program evaluation using retrospective self-assessments and independent performance ratings. *Teach Learn Med.* 2001;13:153-60.
14. Ullian JA, Bland CJ, Simpson DE. An alternative approach to defining the role of the clinician teacher. *Acad Med.* 1994;64:832-8.
15. Baldwin CD, Levine HG, McCormick DP. Meeting the faculty development needs of generalist physicians in academia. *Acad Med.* 1991;155:962-7.
16. Beasley BW, Wright SM, Cofrancesco J Jr, Babbott SF, Thomas PA, Bass EB. Promotion criteria for clinician-educators in the United States and Canada: a survey of promotion committee chairpersons. *JAMA.* 1997;278:723-8.
17. 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;69:152-4.
18. Solomon DJ, Speer AJ, Rosebraugh CJ, DiPette DJ. The reliability of medical student ratings of clinical teaching. *Eval Health Prof.* 1997;20:343-52.