Self-Evaluation by Residents in Family Medicine

Marian R. Stuart, PhD, Harris S. Goldstein, MD, DMedSc, and Frank C. Snope, MD
Piscataway, New Jersey

Accurate self-evaluation is central to a family physician's professional growth both during and after the residency training period. Self-evaluations of residents' interviewing skills, as demonstrated in videotaped interviews with simulated patients, were compared to multiple faculty evaluations as part of an annual assessment. Means for resident evaluations were lower and showed greater variations than faculty ratings but correlated significantly in several areas. Inter-rater reliability coefficients were highest when criteria were most specific. Residents benefited from the opportunity to compare and discuss their perceptions with those of objective and competent raters.

Graduate education in the specialty of family medicine has as its objective the fostering of a broad base of medical knowledge, psychosocial awareness, and effective interpersonal skills. Central to the physician's continued post-residency growth is the acquisition of the ability for accurate and objective self-evaluation.

The Quality Assurance Program of the Department of Family Medicine, CMDNJ-Rutgers Medical School, supported by a grant from the Robert Wood Johnson Foundation, has for three years sought to monitor the quality of the educational experience of residents in its three affiliated residency programs.[1] As part of this program, clinical skills, as demonstrated in interviewing, by tests of cognitive knowledge, and by patient chart recording, were formally evaluated each year. This assessment was intended as a formative evaluation with provision for remediation where necessary. In contrast to a terminal evaluation at the end of the residency, a formative evaluation requires a system of ongoing assessment of resident strengths and weaknesses throughout the residency period with the opportunity for appropriate remediation.

To achieve such an evaluation system, procedures were developed for the assessment of clinical skills through the use of videotaped interviews of residents with simulated patients. The development of criteria based assessment forms was also accomplished. However, it soon became evident that there was a critical need for a sustained high level of resident involvement in the evaluation procedure if it were to be successful.

In previous years, a formal assessment of the resident's skills was made by designated reviewers, with the exact format for feedback to the resident varying somewhat from year to year. In 1976, individual feedback was provided to the residents in the form of percentage ratings of the behaviors they exhibited. An extensive standardized checklist was used which specified the optimal performance for various segments of the simulated patient

From the Departments of Family Medicine and Psychiatry, College of Medicine and Dentistry of New Jersey, Rutgers Medical School, Piscataway, New Jersey. Requests for reprints should be addressed to Dr. Marian R. Stuart, Department of Family Medicine, CMDNJ-Rutgers Medical School, Piscataway, NJ 08854.

0094-3509/80/040639-04$01.00

interview. In 1977, audiofeedback was spliced into the videotape by voice overlay, providing faculty comments as the residents reviewed the tape of the interview. Residents also received a written global evaluation of their skills which pointed out specific instances of exceptionally good or poor interviewing techniques. These methods were time consuming and depended upon the residents actively availing themselves of the time to review the tapes and passively absorbing the feedback.

A reassessment of the procedure in 1978 led to the conclusion that the skills of self-perception and self-evaluation were critically important components in the evolution of both the resident and the future family physician. It was hypothesized that residents able to objectively assess and monitor their own performance have the maximum opportunity to initiate changes to improve their skill level. Self-evaluation and the opportunity to compare ratings with other observers were consequently built into the annual assessment of all residents to promote the development of skill in self-assessment.

In recent years reports in the literature have documented the efficacy of self-observation, self-recording, and self-evaluation in promoting improved performance, feelings of psychological well being, and professional development.[2-6] Experimental studies have shown that self-observation and -evaluation are most effective when desirable behavior is clearly specified, standards and goals are provided, and feedback on the accuracy of the self-evaluation is made available.[7] Several innovative clinical programs have demonstrated that self-evaluation can be successfully utilized.[8-10] In family practice, results of a national survey indicate that 67 percent of residency programs teaching interpersonal skills give some instruction in self-assessment, yet of these only 13 percent report having prepared any self-evaluation materials for residents.[11] Consequently, a decision was made to incorporate formalized self-assessment into our ongoing evaluation program using an evaluation instrument designed to clearly specify performance requirements.

Method

Small group meetings were held several weeks prior to the annual videotaping of interviews with simulated patients. Residents were provided with a sample of the evaluation form to be used which clearly specified all behaviors which were to be demonstrated in order to obtain a criterion performance score (ie, a score indicating full skill acquisition) in each area. Discussion was encouraged. Areas to be evaluated included: (1) the opening phase consisting of preparatory chart review, introduction to the patient, and putting the patient at ease; (2) communication and interviewing skills stressing a variety of components, including control of the interview; (3) appropriateness of the physician's vocabulary and ability to explain clearly; (4) closure specifying a number of required actions; (5) background investigations to include the psychosocial aspect of the patient's situation and family background; (6) response to the patient; (7) maintaining a professional manner; (8) the various aspects of therapy and disposition.

Fifty-six residents in the three years of training (19 R-1s, 17 R-2s, and 20 R-3s) took part in the evaluation. Each resident had two interviews with two different simulated patients. Following these videotaped interviews, the residents were given an opportunity to review their tapes and score them. A faculty member also reviewed and scored each tape. Each interview was scored separately. The resident and the faculty member then discussed the areas of agreement and disagreement in their respective judgments of strengths and weaknesses.

Besides the individual resident and this initial faculty member, two other faculty members also rated each tape. This relatively large number of raters was in keeping with procedures of the previous years and was maintained to keep the data collection and evaluation comparable from year to year. The simulated patients did not feel competent to evaluate the resident's clinical skills but provided in-depth comments on their reactions during the interview.

After all the scores had been compiled, residents met individually with the director of the Quality Assurance Program. At this meeting, the residents were able to compare and discuss their self-evaluations with the evaluations, perceptions, and comments made by all reviewers.

Residents' self-evaluations were compared to the average of the three faculty raters (the preceptor in the individual residency program and two reviewers from the CMDNJ-Rutgers Medical School, Department of Family Medicine).

Table 1. Comparison Between Self and Faculty Assessments of Residents' Clinical Skills In Two Simulated Patient Interviews

Skill Area	Self Score		Faculty Score		
	Mean	Standard Deviation	Mean	Standard Deviation	Correlation Between Assessments†
1. Opening Phase	83.8	12.9	88.1	7.9	0.15
2. Interviewing Techniques	69.9	18.4	76.1	12.4	0.08
3. Vocabulary and Explanations	74.7	15.9	82.2	12.0	0.16
4. Closure	69.9	18.4	73.6	11.7	0.31**
5. Investigations	69.0	14.0	69.9	13.3	0.18
6. Response to Patient	74.4	14.9	78.5	10.9	0.39***
7. Professional Manner	77.1	14.1	88.2	9.8	0.21
8. Therapy and Disposition	70.1	12.8	75.4	10.8	0.24*
Totals	73.6	12.0	79.0	8.7	0.31**

N=56
*P<0.05
**P<0.01
***P<0.001
†Pearson Product Moment Coefficient

Results

Table 1 shows the self and faculty ratings for the 56 residents. Faculty ratings represent the mean of the scores generated by the preceptor and the two Rutgers Medical School raters. It can be seen that the means for the self-evaluations are lower and the variation greater than for the evaluations by the faculty raters. However, the variation of the faculty scores is less because it represents an average derived from three raters. The ratings of the residents and the faculty showed modest but significant correlations in three of the eight skill areas; closure ($r=.31$, $P<.01$); response to patient ($r=.39$, $P<.001$); and therapy and disposition ($r=.24$, $P<.05$). A correlation of .31 ($P<.01$) was obtained when the total score (sum of the eight areas) for the two interviews was considered.

The inter-rater reliability of the program preceptor with the two Rutgers Medical School raters was slightly higher: $r=.375$ ($P<.005$). The reliability coefficients for the two Rutgers Medical School reviewers were .54 and .58, respectively, when each person's rating was correlated with the combined score of the program preceptor and the other Rutgers Medical School rater. These correlat are highly significant ($P<.001$). However, it i portant to note that even the highest i reliability indicates that only 35 percent of the iability of reviewer scores represents agreem t between the raters.

Discussion

The correlation between the residents' self-ratings and the average of the faculty scores may be lower than desired for the individual skills. In skill areas where there is little ambiguity about what is really correct, ie, closure, response to patient, and therapy and disposition, agreement is greatest and the indication for teaching is obvious. There is a need to develop clear and detailed criteria for use in judging performance in behavioral terms. The higher reliability between the Rutgers Medical School raters who shared more specific criteria for performance underscores this point. In the area of interviewing techniques, correlation is particularly low between resident and

faculty ratings. This may indicate that residents and faculty do not share a common conception of what constitutes a good performance. Although criteria for good closure are specific and correlation between raters is high, this is an area where residents see themselves performing at less than 70 percent of criterion (Table 1).

The inter-rater reliability remains a critical area of concern. Raters individually report that they tend to be inconsistent, at one time scoring an item low, and then after observing several residents exhibit the same flaws, tending to record "adequate" for a similar level of performance. Furthermore, raters having previous contact with a particular resident may have a preconceived notion of how that resident performs, rather than attending to the actual interview. Some raters also tend to make errors of leniency by rating first year residents higher than performance warrants. Raters need to be cautioned about these natural tendencies and encouraged to make necessary adjustments.

In-depth interviews with residents after scoring revealed that they succumb to similar types of evaluation errors as do faculty. They tend to rate themselves in line with their assessment of how they usually do, faulting themselves for inconsequential details or not having lived up to unrealistic expectations. Several of the most competent residents were uncomfortable scoring themselves as reaching criterion levels on every item, even though their performance was at that level.

A close examination of the distribution of individual scores indicated that one group of adequately performing residents tended to judge themselves as performing lower than faculty assessments, while another group of less accomplished residents tended to rate themselves much better than their actual performance. Importantly, there were many instances where residents were scoring themselves as having done well when actually there were serious omissions. Here the need for remediation is clear. When these inconsistencies were graphically pointed out, residents were often surprised. They appeared genuinely open to this type of information, raising their awareness level and their ability to make necessary adjustments.

In view of the above, the setting of reasonable and discernable criteria is of the utmost importance. It would be ideal if there were accurate (valid) and consistent (reliable) measures of interpersonal skills. Since for the physician in practice there is no preceptor to make judgments, the perceptions of the individual physicians become all important. Their measurements (judgments) will be subject to a variety of human errors, but the continued attempt to define and measure as accurately as possible must be encouraged. Ongoing self-monitoring should be fostered in order to maximize performance and professional satisfaction.

The opportunity for residents to discuss their self-ratings and compare them to those of objective and respected evaluators, challenging when appropriate, becomes an intense and meaningful learning experience. This experience will help residents to become more accurate and objective self-observers, in a better position to assure continuing quality of care for their patients.

Acknowledgements

The Quality Assurance Program is supported by a grant from the Robert Wood Johnson Foundation.

References

1. Sadler GR, Snope FC, Currie BF: A quality assurance program for graduate education in family medicine. J Fam Pract 4:751, 1977
2. Adams WR, Ham TH, Mawardi BH, et al: Research in self-evaluation for clinical teachers. J Med Educ 49:1166, 1974
3. Amatora M Sr: Validity in self-evaluation. Educ Psychol Measur 16:119, 1956
4. Mahoney MJ, Moore BS, Wade TC, et al: Effects of continuous and intermittent self-monitoring on academic behavior. J Consult Clin Psychol 41:65, 1973
5. Marriner A: Student self-evaluation and contracted grade. Nurs Forum 13(2):130, 1974
6. Morgan MK, Irby DI: Evaluation of Clinical Competence in the Health Professions. St. Louis, CV Mosby, 1978
7. Kazdin AE: Reactive self-monitoring: The effects of response desirability, goal setting, and feedback. J Consult Clin Psychol 42:704, 1974
8. Abrams RG, Kelly ML: Student self-evaluation in a pediatric operative technique course. J Dent Educ 38:385, 1974
9. Geissler PR: Student self-assessment in dental technology. J Dent Educ 37(9):19, 1973
10. Kennell JH, Tempio CR, Wile MA: Self-evaluation by first year medical students in a clinical sciences programme. Br J Med Educ 7(4):230, 1973
11. Kahn G, Cohen B, Jason H: Teaching interpersonal skills in family practice: Results of a national survey. J Fam Pract 8:309, 1979