Development of Reflective Practicioners: *Comparison of Obstetrics and Gynecology Resident Self-Ratings to Faculty Ratings*

Karla Hemesath, PhD* • Raj Narayan, MD, FACOG, FRCOG**

*Department of Obstetrics and Gynecology, University of Wisconsin Medical School-Milwaukee Clinical Campus, and the Aurora Health Care Obstetrics and Gynecology Residency Program, Milwaukee, Wisconsin

**Department of Obstetrics and Gynecology, Medical College of Wisconsin, Milwaukee, Wisconsin. Former Associate Program Director, Aurora Obstetrics and Gynecology Residency Program, Milwaukee, Wisconsin

Abstract

Purpose

The purpose of this project was to compare resident self-assessments of performance to faculty ratings of performance. Developing proficiency in self-assessment is an important task for residents and for life-long learning and performance improvement (DeAngelis, 2003). If relative congruence exists between one's self-perception and the perceptions of others, the ongoing process of hearing and assimilating feedback and acquiring knowledge is more effective and less painful for all involved. To assist our residents in the development of self-assessment skills we began requiring them to complete a self-evaluation that addressed performance in each of the six core competencies.

Methodology

In the 2004-2005 academic year, each of our eight residents completed the self-assessment form before their semi-annual evaluations (January and June). The self-assessment tool is a modified version of the CREOG Task Force's "Obstetrics and Gynecology Resident Evaluation Form"[1]. The form assesses resident performance in each of the six competencies (patient care, medical knowledge, professionalism, communication skills, practice-based learning and improvement, and systems-based practice). Patient care is split into clinical skills and surgical skills and we added a category assessing resident teaching skills because our residents spend significant time interacting with medical students. Resident performance is assessed using a 9 point scale with responses 1-3 grouped in the "unsatisfactory" category, responses 4-6 grouped in the "satisfactory" category and responses 7-9 grouped in the "excellent" category. Performance descriptors are provided for each category of responses.

The program faculty members completes the "Obstetrics and Gynecology Resident Evaluation Form" for each resident in advance of the semi-annual evaluation process conducted in January and June.

The faculty data for each resident was compiled for both semi-annual evaluations and trends across raters were examined. The averaged faculty ratings were then compared to the resident's own ratings. Written comments from the faculty evaluations were analyzed for common themes and agreement and then compared to resident comments.

Results

Faculty ratings in both fall and spring were similar in except in the case of one resident. In comparing resident self assessment to faculty assessment during the fall evaluation period, two residents' self-assessments were significantly lower than those of faculty while the rest of the residents' self-assessments were not significantly different that the faculty assessments. Greater differences between resident and faculty assessments (both lower and higher) were found in the spring evaluation period. Differences were found in the overall faculty assessment of resident performance in both fall and spring evaluation periods with one resident's performance rated higher than the others and one residents' performance significantly lower.

Condusions

Comparison of self-assessment to faculty assessment in both evaluation periods yielded mixed results. Some residents rated their performance higher than did faculty, some lower than the faculty and others about the same. Regardless, we believe that development of effective self-assessment skills is critical. Self-assessment forms the basis for practice-based learning and improvement and we will continue to require residents' to self assess as a means to foster continued self awareness, personal and professional growth.

Introduction

Our main objective was to compare resident self-assessment of performance in the ACGME core competencies to faculty assessment of resident performance. By attending faculty using the Obstetrics and Gynecology Resident Evaluation Form (CREOG Taskforce, 2003).

- Consists of nine questions to assess aspects of professionalism including communication with patients and ancillary staff, respect for patients and ancillary staff, compassion, advocacy, reliability, responsibility, and honesty/integrity.
- Form was modified by adding a question that assessed resident teaching skills and space for goal setting and comments.
- Respondents use a nine point Likert scale to assess performance on each of the nine items and an option "unable to evaluate" was included.
- Responses 1-3 fall into the unsatisfactory category, responses 4-6 into the satisfactory category and responses 7-9 into the excellent category.
- Brief descriptions of exemplary behavior are provided for each question, but descriptors of behavior are not provided for each category so respondents must personally decide what they consider to be unsatisfactory, satisfactory and excellent.

Methodology

- During the 2004-2005 academic year eight residents were in training (3 OBG1; 2 OBG2, 2 OBG3).
- Attending faculty completed the Obstetrics and Gynecology Resident Evaluation Form (CREOG Taskforce, 2003) electronically through the E*Value system (Advanced Informatics, LLC, Minneapolis, MN).
- At the two semi-annual evaluations (November and April) residents self-assess using the same form that the faculty use with additional space for free text, goal setting and comments.
- Each resident received between 18-2 completed faculty semi-annual evaluations.
- Average scores were calculated for each assessment because some respondents indicated that they were unable to assess the resident on questions.
- To compare nurse and peer scores to faculty assessments of professionalism, nursing and peer scores were correlated to global faculty assessments of resident professionalism (item #5 in the Obstetrics and Gynecology Resident Evaluation Form) completed for each semi-annual evaluation.
- There were an average of 21 (range=18 to 24) faculty assessments for each resident.

Results

Table 1: Faculty Ratings

Resident	Fall Average	Range Average	Spring	Range	Diff*	P value
A	7.28^{a}	5.4-9	7.43^{a}	4.57-9	.15	.293
B	5.39^{c}	4.5-6.43	6.14^{b}	4-6.5	.75	0.03
C	6.77^{b}	6.3-8.4	6.71^{d}	5.75-8	-.06	.439
D	5.83^{b}	5.75-7	6.28^{c}	4.75-7.25	.45	.148
E	5.57^{c}	4.5-6.43	5.89^{c}	4-7	.32	.217
F	5.80^{c}	4.13-7	5.39^{d}	4.125-7	-.41	.165
G	6.09^{b}	4.5-7.5	6.15^{c}	4.625-7.38	.06	.436
H	4.69^{d}	2.57-6.8	4.99^{e}	3.86-6.43	.30	.258

**Difference between fall faculty average rating and spring faculty average scores.*

Note: Means with the same superscript are not significantly different at p <. 05

- In all cases but one (resident B), faculty scores did not differ significantly between fall and spring.
- Faculty rated all residents but two (residents C and F) higher in the spring than in the fall, but the difference was not statistically significant.
- When faculty ratings were compared across resident ratees using analysis of variance (ANOVA) the means were significantly different in both fall and spring.
- Using Fisher's Least significant difference (LSD) test, one resident (resident A) was rated significantly higher than peers in both fall and spring and one residents (resident H) was rated significantly lower than peers in both fall and spring.

Table 2: Self Ratings

Resident	Fall Average	Spring Average	Difference	P value
A	5.43	5.57	0.14	0.369
B	5.88	6.25	0.37	0.249
C	4.86	5.25	0.39	0.163
D	6.63	6	-0.63	0.157
E	5.43	5	-0.48	0.169
F	6.25	6.375	.125	0.347
G	5.5	5.5	0	
H	5.00	5.625	0.625	0.006

- Most residents rated themselves higher in the spring than in the fall with the exception of two (residents D and E).
- Only one resident's self assessment (resident H) was statistically higher in the spring than in the fall.

Table 3: Comparison of Faculty and Self Ratings

Resident	Faculty Fall	Self Fall	Diff	Faculty Spring	Self Spring	Diff
A	7.28	5.43	1.85 (p =.0002)	7.43	5.57	1.86 (p=.0005)
B	5.39	5.88	-.49 (p=.162)	6.14	6.25	-0.11 (p=.405)
C	6.77	4.86	1.91 (p=.0003)	6.71	5.25	1.46 (p=.00002)
D	5.83	6.625	-.80 (p=.103)	6.28	6.00	0.28 (p=.256)
E	5.57	5.43	0.14 (p=.359)	5.90	5.00	0.90 (p=.024)
F	5.80	6.25	-.045 (p=.106)	5.39	6.38	-0.99 (p=.0118)
G	6.09	5.5	0.59 (p=.134)	6.15	5.5	0.65 (p=0.03)
H	4.70	5.00	-.30 (p=.213)	4.99	5.625	-0.64 (p=0.035)

- In the fall, only two residents' self-assessment scores were significantly different (lower) than the faculty ratings.
- In the spring, most residents except two (residents B and D) self-assessment scores were statistically different than those of the faculty.

Differences among raters

- To assess the performance of the faculty raters, ANOVAs were conducted to assess if the means for each rater were significantly different.
- In both fall and spring, the mean scores across faculty raters were significantly different. Fisher's LSD test indicated that one faculty rater consistently rated residents higher than the others and another faculty member consistently rated residents lower than the other faculty raters.

Discussion

- Faculty ratings between fall and spring evaluation periods were consistent with only one resident rated significantly higher in spring than in fall. In one respect this seems reasonable because it indicates a level of consistency. Alternatively, one would expect that residents' performances should have improved over the course of six months. This should particularly be true in the case of PGY1 residents.
- There was some difference between resident self-assessment and faculty assessment in both fall and spring, but more residents' scores differed from faculty in the spring evaluation.
- Assessment skills must be taught to faculty and to residents. Rater education is critical. It is important to formally acquaint the assessors with not only the tool that they are to use, but to explain the context for the assessment and how the information will be used.
- When faculty raters' mean scores were compared for both evaluation periods, significant differences were found. In both evaluation periods, one resident's mean evaluation score was significantly higher than the other residents and one resident's mean score was significantly lower than the other residents.
- When mean scores were compared across raters for both evaluation periods, significant differences were found. One faculty member's mean rating was higher than the others and one faculty member's mean rating was lower than the others.
- These differences among raters lead us to suspect bias that Goldman (1996) refers to the "halo" effect (positive) or the "millstone" effect (negative).
- Davis (2002) found that unlike comparison of attending and peer evaluations, there was no correlation between attending and self-evaluations. He argued that the residents were very good at evaluating other residents but lack adequate self-assessment skills.
- Risucci et al (1989) found that self-ratings were significantly higher than supervisor ratings on judgment and reliability.

Conclusions

- Comparison of self-assessment to faculty assessment in both evaluation periods yielded mixed results. Some residents rated their performance higher than did faculty, some lower than the faculty and others about the same.
- Despite our mixed findings, we agree with the conclusion that (Davis, 2002) asserts in that self-evaluation by residents provides the opportunity to develop or improve their ability to honestly appraise their skills.
- Development of effective self-assessment skills is critical. Self-assessment forms the basis for practice-based learning and improvement and we will continue to require residents' to self assess as a means to foster continued self awareness, personal and professional growth.
- DeAngelis (2003) also discusses the importance of accurate self-assessment in the development of clinical reasoning and judgment skills.

References

CREOG Competency Task Force, (2003). Professional Associate Questionnaire.

Davis J.D. (2002) Comparison of faculty, peer, self and nurse assessment of obstetrics and gynecology residents. Obstetrics and Gynecology, 99:647-651.

DeAngelis, T. (2003). Better self-assessment equals better medical judgments. APA Monitor. 34(2).

Di Matteo M, R. and DiNicola D.D. (1981) Sources of assessment of physician performance: A study of comparative reliability and patterns of inter-correlation. Medical Care 19: 829-842.

Goldman R.L. and Ciesco E. (1996) Improving peer review: alternatives to unstructured judgments by a single reviewer. Joint Commission Journal of Quality Improvement, 22:762-769.

Risucci D.A., Tortolani A.J. and Ward R.J. (1989) Ratings of surgical residents by self, supervisors and peers. Surg Gynecology Obstet 169:519-526.