Multirater Assessment of Resident Professionalism

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

*Department of Obstetrics and Gynecology, University of Wisconsin Medical School and Aurora 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

As part of our program's revision of resident evaluation process, we implemented use of the Professional Associate Questionnaire (CREOG Competency Task Force, 2003) to gather evaluation data on aspects of resident performance from resident peers and nursing staff. This provides a range of data for formative feedback to residents and identifies areas of resident strength and opportunities for performance improvement. The main purpose of this study is to determine if statistically significant differences exist between staff ratings of resident professionalism and peer assessment. With the ongoing implementation of the evaluation of ACGME Core Competencies, we are interested in the effectiveness of Professional Associate Questionnaire as a tool for multi-rater assessment of resident professionalism.

Methods

The Professional Associate Questionnaire was developed by the CREOG Competency task force[2] through modification of the American Board of Internal Medicine (ABIM)'s Peer Assessment questionnaire. The 9-item Professional Associate Questionnaire assesses resident performance in the areas of communication, respectfulness, compassion, reliability, honesty/integrity, responsibility and patient advocacy.

Residents were required to complete the Professional Associate Questionnaire for each of their resident peers in preparation for our semi-annual evaluation process. The Professional Associate Questionnaire was sent to each resident electronically through our online evaluation system (E*Value, Advanced Informatics, Minneapolis, MN). Residents completed one for each of their 7 peers in our program. Additionally, nursing staff members in the inpatient units, labor and delivery and Resident Continuity Clinic (RCC) were also asked to complete the Professional Associate Questionnaire. The attending faculty completed the Obstetrics and Gynecology Resident Evaluation Form (CREOG Taskforce, 2003) and for the purposes of this study, we used only the scores from the professionalism question (item # 5).

Statistical analysis was conducted using SAS (SAS Institute, SAS Circle, Cary, NC).

Results

Each resident received an average of 23 completed evaluations from resident peers and nursing staff in the 2004-2005 academic year (range 21-27). A total of 194 completed questionnaires were analyzed. There was a significant difference between nurse and peer ratings for PGY1 and PGY2 residents and not for PGY3 residents. For each resident the difference was not statistically significant. When the questions were taken individually, there was no difference between nurse and peer ratings. There was positive correlation between faculty, nurse and peer ratings supporting the hypothesis that the Professional Associate Questionnaire is measuring the same qualities as the professionalism part of the Global rating form.

Conclusions

The Professional Associate Questionnaire is a useful method of assessing resident professionalism across groups of raters. Future studies include item assessment and performance analysis, content validity studies and external validity assessment. Given its short, concise form it is a useful and practical method to obtain formative data on resident professionalism.

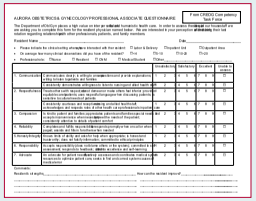
Background

- The American Board of Internal Medicine's (ABIM) Charter on Medical Professionalism[1] states that "professionalism is the basis of medicine's contract with society".
- Papadakis and colleagues [3] reviewed 68 cases arbitrated by the California Medical Board and found that 95% of these cases centered on acts of unprofessional behavior. During the period of this study, a disproportionately higher number of obstetrician-gynecologists were represented among those disciplined by the Board.
- Arnold [4] commented that "without solid assessment tools, questions about the efficacy of approaches to educating learners about professional behavior will not be effectively answered".
- Lynch [5] noted in her extensive review of assessment tools it is necessary to study the differences between various raters' view points when a particular tool is used to evaluate residents and to interpret it in the context of the stage of the learner's development.
- Duff [6] outlined humility, honesty, responsibility, reliability and accountability as major characteristics of professionalism in medical students, residents and practicing physicians.
- The American Board of Internal Medicine (ABIM) Professional Associate Rating Form was developed.
- The Council for Resident Education in Obstetrics and Gynecology (CREOG) Competency Taskforce adapted this to develop the Professional Associate Questionnaire for the assessment of professionalism in residents in Obstetrics and Gynecology.

Introduction

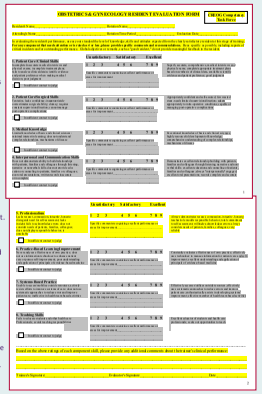
Our main objective was to study the use of Professional Associate Questionnaire in the assessment of Obstetrics and Gynecology residents' professionalism by nurses and peers and correlate the result with the assessment by attending faculty using the Obstetrics and Gynecology Resident Evaluation Form (CREOG Taskforce, 2003).

- Professional Assessment Questionnaire 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.
- 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).
- At the two semi-annual evaluations (November and April) residents were required to complete the Professional Associate Questionnaire for each peer electronically through the E*Value system (Advanced Informatics, LLC, Minneapolis, MN).
- Nursing staff from labor and delivery, the inpatient units, resident continuity clinic and nurse midwives were asked to complete the Professional Associate Questionnaire in paper form with results entered by the program office staff.
- Each resident received between 21-27 completed Professional Associate Questionnaires.
- 194 total evaluations collected in both evaluation periods (97 nursing evaluations and 97 peer evaluations).
- Average scores were calculated for each assessment because some respondents indicated that they were unable to assess the resident on questions.
- Attending faculty completed the Obstetrics and Gynecology Resident Evaluation Form (CREOG Taskforce, 2003) and for the purposes of this study, we used only the scores from the professionalism question (item # 5).
- Attending faculty members were unaware of the nurse and peer evaluations that were sent directly to the Program Director.
- 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.
- To compare nurse and peer scores to self-assessments of professionalism, nursing and peer scores were correlated to global self-assessments of resident professionalism (item #5 in the Obstetrics and Gynecology Resident Evaluation Form) completed for each semi-annual evaluation.
- Residents self-assess using the same form that the faculty use with additional space for free text, goal setting and comments.



Results

Table 1: Peer Evaluation Scores

Resident	Fall Mean Score	Spring Mean Score	Score Change	P value
A	7.6458	7.042	-.6038	p=.0614
B	6.751	7.368	.617	p=.1378
C	7.87	7.46	-.41	p=.2542
D	7.842	7.257	-.585	p=.1557
E	7.324	7.655	.331	p=.3037
F	6.671	6.409	-.262	p=.3688
G	7.578	7.421	-.157	p=.422
H	6.64	6.558	-.082	p=.4559

In all but two cases, mean peer ratings were lower in the spring evaluation than the previous fall ratings though none of the differences were statistically significant.

When comparing the peer mean scores across residents through analysis of variance (ANOVA), the differences in either fall or spring were not statistically significant.

Table 2: Nurse Evaluation Scores

Resident	Fall Mean Score	Spring Mean Score	Difference	P value
A	7.65	7.47	-.18	.42
B	7.72	7.53	-.19	.40
C	7.98	8.15	0.17	.357
D	7.56	7.87	.31	.29
E	7.86	7.49	-.37	.321
F	6.82	7.11	.29	.377
G	7.10	7.33	.23	.374
H	7.01	6.32	-.69	.27

In four cases, mean nurse ratings were lower in the spring evaluation than the previous fall ratings. None of the differences were statistically significant.

When comparing the nursing mean rating scores across residents through analysis of variance (ANOVA), none of the differences among the scores were statistically significant in either fall or spring.

Table 3: Peer and Nurse Score Comparison By Resident Class

	PGY1 (n=3)		PGY2 (n=2)		PGY3 (n=3)	
	Peer	Nursing	Peer	Nursing	Peer	Nursing
Number of evaluations	33	34	25	24	39	39
Score Range	3-9	3-9	3-9	3-9	3-9	3-9
Mean	7.06	7.4	6.93	7.28	7.584	7.644
Standard Deviation	1.30	1.54	1.293	1.567	1.249	1.31
Difference	Significant (p = 0.024)		Significant (p=.00013)		Not Significant (p=0.199)	

The average score within each group of raters was calculated for each resident and the resident averages combined to obtain a class average for each of the three classes of residents. We did not have any residents in PGY4 during the study period.

Peer mean scores were compared to nursing scores for each class using Student's ttest.. T-test differences were considered statistically significant with P < 0.05.

Significant differences were found between nurse and peer ratings for Postgraduate years 1 and 2 but not for Year 3. The nurses' scores were higher than that of the peers on all the questions (Table 3).

Table 4: Question Score Comparisons

	Nursing		Peer		
Question Topic	Mean	Range	Mean	Range	Difference
Communication with Patients	7.22	4-9	7.05	4-9	n.s.
Communication with Staff	7.40	4-9	7.02	5-9	n.s.
Respectfulness to Staff	7.46	3-9	7.07	5-9	n.s.
Respectfulness to Patients	7.60	4-9	7.31	3-9	n.s.
Compassion	7.54	3-9	7.38	5-9	n.s.
Reliability	7.41	3-9	7.34	3-9	n.s.
Honesty/Integrity	7.67	4-9	7.36	3-9	n.s.
Advocate	7.41	3-9	7.28	4-9	n.s.
Responsibility	7.43	3-9	7.28	4-9	n.s.

No significant differences existed between nursing scores on individual questions and peer questions, although nurses rated residents higher than the residents' peers on all questions.

Differences among raters

- To assess the performance of the raters, ANOVAs were conducted to assess if the means for each rater were significantly different.
- In both fall and spring, the mean scores across peer raters were significantly different. Fisher's LSD test indicated that one peer rater consistently rated peers higher than the others and another peer rater consistently rated peers lower than the others.
- The same process was conducted with the data from the nursing raters and the significant differences in mean nursing rating were found across the nursing raters.

Table 5: Relationship of Peer and Nursing Ratings to Other Assessments of Professionalism

Resident	Peer	Nurse	Faculty*	Self*
A	7.34	7.56	7.46	6
B	7.06	7.65	6.32	7
C	7.67	8.07	7.33	6
D	7.55	7.72	6.73	8
E	7.49	7.68	6.45	6
F	6.54	6.97	6.14	7
G	7.50	7.22	6.48	5.5
H	6.60	6.67	5.57	5.5

*Scores were obtained from the professionalism question on the faculty semi-annual evaluations and the semi-annual resident self-assessment questionnaires

- In all but one case, the faculty rated residents lower than peers.
- In all cases, nurses rated residents higher than their peers.
- In all cases nurses rated residents higher than the faculty.
- Two residents rated themselves higher on professionalism than did nursing staff, peers and faculty.

Table 6: Score Correlations

	Faculty Assessment	Self Assessment
Peer Ratings	r =.783	r =-.002
Nurse Ratings	r =.736	r =.282

- Peer scores were compared separately with both faculty and self-assessment of professionalism through calculation of the Pearson correlation coefficient.
- Nurse scores were compared separately with both faculty and self-assessment of professionalism through calculation of the Pearson correlation coefficient.
- Both peer and nurse ratings of professionalism were strongly positively correlated with faculty assessment of resident professionalism.
- Nurse assessment of professionalism was also positively correlated with self-assessment of professionalism.
- No linear relationship was found between self-assessment of professionalism and peer assessment of professionalism.

Discussion

- Ramsey and Wenrich[7] concluded that the use of professional associate ratings (PAR) is a valuable assessment measure for assessing the performance of practicing physicians. Their study indicated that the physician being evaluated could select his/her raters without significantly biasing results and that approximately 10-12 ratings from peers provide a reliable rating of an individual physician.
- Ramsey and Wenrich's findings also suggested that ratings from approximately 10-15 nurse raters provide a reliable rating of individual physicians. They also noted that ratings completed by nurses showed lower ratings in several humanistic categories including respect, integrity and responsibility in contrast to medical knowledge and verbal communications when compared to physician peer ratings.
- Kaplan and Centor[8] reported on the validity of nurse evaluation of residents' humanistic behavior and found that nurses were able to discriminate between various levels of behavior and these were more broadly distributed than physician ratings.
- Ramsey and Wenrich[7] also noted that there has been consistent findings across studies that nurses appear able to evaluate residents reliably and with good discrimination and relatively small number of nurse ratings is required to obtain a reliable rating. This contrasted with assessment by attending faculty who provide a more single-dimensional rating.
- In a study comparing faculty, peer, self and nurse assessment of Obstetrics and Gynecology residents (Davis, 2002)[9] found a poor reliability of the nurse ratings when evaluating clinical competency and interpersonal skills. They did not use the Professional Associate Questionnaire and the major humanistic qualities were not assessed individually.
- One of the greatest advantages to use of a professional assessment rating (PAR) such as the Professional Associate Questionnaire is the ability to use the data collected for formative performance feedback (Ramsey and Wenrich, 1999)[7].
- While the validity and reliability of measures such as this are not definitively established to support their use in summative, high stakes evaluation, they do provide extremely valuable information for performance maintenance and improvement.
- 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.
- As we investigated original evaluation forms for each resident, we encountered numerous examples of what we term "serial 7's, serial, 6's or serial 5's" in which the respondent would choose the same response for each question.
- Assessments, when performed individually, are subject to bias, either positively (the "halo" effect) or negatively (the "millstone" effect) and is often "hit or miss" in terms of random observations (Goldman, 1996)[10].
- Significant positive correlation between the global professionalism question completed by faculty and the nurse and peer evaluation of professionalism using the Professional Associate Questionnaire. This supports the hypothesis that the Professional Associate Questionnaire is measuring the same qualities as the professionalism part of the Global rating form.
- Multi-rater feedback is more meaningful and provides more useful data when intervention is needed.

Study Limitations

- Data is from one academic year and limits generalizability of our results.
- Several residents did not complete peer assessments for each resident and the peer data set was not complete.
- Nursing data was submitted to the program office in paper form and rater identification was not maintained as the data was entered into the evaluation program.

Conclusions

- The Professional Associate Questionnaire provides useful data from peers and nursing staff for resident feedback and formative assessment.
- While its psychometric properties have not been fully addressed, it seems to be a reasonable tool to use in the 360-degree evaluation of resident professionalism.
- Peer, faculty and nursing assessments of professionalism are not correlated with self assessment of professionalism.
- The results were consistent between resident peers, nursing staff and faculty.
- Verbatim comments from the assessments are particularly powerful anecdotes to reinforce positive traits and to provide information for resident growth and skill development.

References

1. CREOG Competency Task Force, (2003). Professional Associate Questionnaire.
2. ABIM (2002) Medical Professionalism in the new millennium: a physician charter. Annals of Internal Medicine, 136:243-246.
3. Papadakis M.A., Hodgson C.S., Teherani A. and Kohatsu N. (2004). Unprofessional behavior in medical school is associated with subsequent disciplinary action by a state medical board. Academic Medicine, 79, 244-249.
4. Arnold L. (2002). Assessing professional behavior: yesterday, today, and tomorrow. Academic Medicine, 77:502-515.
5. Lynch D.C., Surdyk P.M. and Eiser A.L. (2004). Assessing professionalism: a review of the literature. Medical Teacher, 26 (4):366-373.
6. Duff P. (2004). Teaching and assessing professionalism in medicine. Obstetrics and Gynecology, 104: 1362-1366.
7. Ramsey P.G. and Wenrich M.D. (1999). Use of professional associate ratings to assess the performance of practicing physicians: Past, present, future. Advances in Health Science Education, 4: 27-38.
8. Kaplan C.B. and Centor R. M. (1990). The use of nurses to evaluate house-officers' humanistic behavior. Journal of General Internal Medicine, 5:410-414.
9. Davis J.D. (2002). Comparison of faculty, peer, self and nurse assessment of obstetrics and gynecology residents. Obstetrics and Gynecology, 99:647-651 .
10. Goldman R.L. and Cleso E. (1996). Improving peer review: alternatives to unstructured judgments by a single reviewer. Joint Commission Journal on Quality Improvement, 22:762-769 .