

A multi source feedback program for anesthesiologists

[Un programme de rétroaction multisources pour les anesthésiologistes]

Jocelyn M. Lockyer PhD,*† Claudio Violato PhD,* Herta Fidler MSc†

Purpose: To assess the feasibility, validity, and reliability of a multi source feedback program for anesthesiologists.

Methods: Surveys with 11, 19, 29 and 29 items were developed for patients, coworkers, medical colleagues and self, respectively, using five-point scales with an 'unable to assess' category. The items addressed communication skills, professionalism, collegiality, continuing professional development and collaboration. Each anesthesiologist was assessed by eight medical colleagues, eight coworkers, and 30 patients. Feasibility was assessed by response rates for each instrument. Validity was assessed by rating profiles, the percentage of participants unable to assess the physician for each item, and exploratory factor analyses to determine which items grouped together into scales. Cronbach's alpha and generalizability coefficient analyses assessed reliability.

Results: One hundred and eighty-six physicians participated. The mean number and percentage return rate of respondents per physician was 17.7 (56.2%) for patients, 7.8 (95.1%) for coworkers, and 7.8 (94.6%) for medical colleagues. The mean ratings ranged from four to five for each item on each scale. There were relatively few items with high percentages of 'unable to assess'. The factor analyses revealed a two-factor solution for the patient, a two-factor solution for the coworker and a three-factor solution for the medical colleague survey, accounting for at least 70% of the variance. All instruments had a high internal consistency reliability (Cronbach's $\alpha > 0.95$). The generalizability coefficients were 0.65 for patients, 0.56 for coworkers and 0.69 for peers.

Conclusion: It is feasible to develop multi source feedback instruments for anesthesiologists that are valid and reliable.

Objectif: Évaluer la faisabilité, la validité et la fiabilité d'un programme de rétroaction multisources pour les anesthésiologistes.

Méthode: Des sondages comportant 11, 19, 29 et 29 éléments ont été élaborés pour les patients, les collègues de travail, les collègues médecins et nous-mêmes respectivement, en utilisant des échelles en cinq points dont une catégorie «impossible d'évaluer». Les éléments concernaient les habiletés de communications, le professionnalisme, la collégialité, la formation professionnelle continue et la collaboration. Chaque anesthésiologiste était évalué par huit médecins, huit collègues de travail et 30 patients. La faisabilité a été évaluée par les taux de réponses pour chaque instrument. La validité a été évaluée par les profils de cotation, le pourcentage des participants incapables d'évaluer le médecin pour chaque élément et les analyses factorielles exploratrices pour déterminer quels étaient les éléments regroupables avec l'usage d'une échelle. La fiabilité a été évaluée par les analyses du coefficient Alpha de Cronbach et de généralisabilité.

Résultats: On a compté 186 médecins participants. Le nombre moyen de répondants et de pourcentage de questionnaires retournés par médecin a été de 17,7 (56,2 %) pour les patients, 7,8 (95,1 %) pour les collègues de travail et 7,8 (94,6 %) pour les collègues médecins. Les scores moyens étaient de quatre ou cinq pour chaque élément de chaque échelle. Il y a eu relativement peu d'éléments avec de hauts pourcentages de réponses «impossible d'évaluer». Les analyses factorielles ont révélé une solution bifactorielle au sondage des patients, une bifactorielle à celui des collègues de travail et une trifactorielle à celui des collègues médicaux, ce qui constitue au moins 70 % de la variance. Tous les instruments avaient une fiabilité de forte cohérence interne (coefficient α de Cronbach $> 0,95$). Les coefficients de généralisabilité ont été de 0,65 pour les patients, 0,56 pour les collègues de travail et 0,69 pour les pairs.

Conclusion: Il est faisable d'élaborer des instruments valides et fiables de rétroaction multisources pour les anesthésiologistes.

From the Department of Community Health Sciences,* the Faculty of Medicine, University of Calgary; and the Office of Continuing Medical Education,† Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada.

Address correspondence to: Dr. Jocelyn Lockyer, Continuing Medical Education and Professional Development, Faculty of Medicine, University of Calgary, 3330 Hospital Drive N.W., Calgary, Alberta, T2N 4N1, Canada. Phone: 403-220-4248; Fax: 403-270-2330; E-mail: lockyer@ucalgary.ca

Funding: Funding for this work was provided by the College of Physicians and Surgeons of Alberta.

Accepted for publication July 19, 2005.

Revision accepted August 26, 2005.

THE training requirements for anesthesia set by the Royal College of Physicians and Surgeons of Canada stress the attainment of skills as medical expert/decision maker, as well as six other competencies, namely, collaborator, scholar, manager, health advocate, communicator and professional.¹ While it is possible to develop and monitor these competencies in residency educational programs, it is more challenging for practicing physicians to obtain systematic feedback about performance, particularly for non-technical care. Incident reporting,² observational studies^{2,3} simulator-based assessments² and electronic tracking systems⁴ have been successfully adopted as part of quality improvement strategies for anesthesiologists.

In other disciplines of medicine, multi source feedback (MSF) or 360° evaluation is being used to assess and provide feedback to physicians about a broad range of competencies^{5,6} by licensing authorities,⁷⁻⁹ professional organizations^{10,11} and health care facilities.¹² This type of assessment relies on questionnaires completed by patients, medical colleagues, and coworkers to provide feedback to physicians about their communication skills, interpersonal skills, collegiality, medical expertise, and the ability to learn continually and improve practice patterns.⁵⁻¹² There have been a number of studies in which anesthesiologists receive feedback from surgeons¹³ and patients.¹⁴⁻¹⁶ Unlike family medicine,⁷ internal medicine,⁹⁻¹² and surgery,⁸ however, questionnaires from multiple groups (e.g., patients, peers and coworkers) have not been used simultaneously to assess and provide performance data to anesthesiologists.

Studies of MSF show that reliable and valid instruments (questionnaires) can be developed.⁵⁻¹² It appears feasible to develop quality improvement programs in which most physicians within a given discipline can be assessed by eight to ten coworkers and medical colleagues, and 25 patients. This number of raters produces an acceptable reliability for both the overall instrument and the physician being assessed.^{7,9,10-12} Furthermore, given that the intent of MSF is to guide professional development, studies show that participating physicians will use their feedback data to make changes in their practices.^{6,8,10,17}

The practice of anesthesia is clearly different from the practice of a consultant internist, surgeon, or psychiatrist. For same-day admit and surgical day care patients who are not evaluated by an anesthesiologist in a preoperative assessment unit, patient encounters with the anesthesiologist are often brief, and may be compromised by patient stress and/or the effect of anxiolytics administered preoperatively.

The main purpose of the present study was to assess the feasibility, validity, and reliability of a MSF system for the practice of anesthesia. We sought to address the following specific questions: 1) What is the feasibility of an assessment system for practicing anesthesiologists which provides feedback from patients, coworkers, medical colleagues and self? 2) What questions about an anesthesiologist's practice can patients, coworkers, and medical colleagues answer? 3) What are the score profiles for each of the items (i.e., mean, and standard deviation) on the surveys? 4) Do the items on a survey group together into meaningful scales to guide performance improvement direction? 5) Are the instruments reliable for both the practice of anesthesia and for the individual anesthesiologist?

Method

The practice of anesthesia is the most recent component of the College of Physicians and Surgeons of Alberta, Physician Achievement Program (CPSA-PAR) begun in 1996¹⁸ in which medical colleagues, coworkers, and patients provide assessments of physicians in practice. The original goal of the program was to provide feedback to physicians about six broad categories of performance - medical knowledge and skills, attitudes and behavior, professional responsibilities, practice improvement activities, administrative skills, and personal health.⁷ The program is mandatory, with every physician expected to participate every five years. Instruments have been developed and tested for family physicians⁷ and surgeons,⁸ for pediatric, internal medicine and psychiatry specialists⁹ but not for anesthesiologists.

A working group was recruited to develop a set of instruments for these physicians that would be similar to previously developed instruments which had been found to be reliable and valid.⁷⁻⁹ The committee was charged with the mandate of reviewing previous instruments but realigning and redesigning items to assess anesthesia practice. After the working group determined the items for inclusion on the instruments, copies of the instruments were sent to all of the anesthesiologists on the CPSA register for review and feedback as part of the face validation process. The working group made adjustments to the instruments based on this feedback. All previous instruments had included patient surveys and there was lengthy discussion about the advantages and disadvantages of maintaining a patient instrument for anesthesiologists. A small pilot project involving four anesthesiologists indicated that patient responses could be obtained, but would be challenging. In the end, the committee retained an abbreviated patient survey along with

TABLE I Descriptive statistics and factor analysis for patient surveys

Items	<i>n</i>	Mean	S.D.	% U/A	Factors	
					Professionalism	Communication
Q1	3,080	4.62	0.68	1.8	0.82	
Q2	3,100	4.60	0.67	1.1	0.79	
Q3	3,028	4.62	0.67	3.4	0.82	
Q4	3,117	4.75	0.59	0.6	0.86	
Q5	2,530	4.24	1.01	19.3		0.89
Q6	2,192	4.20	1.05	30.1		0.90
Q7	2,797	4.49	0.74	10.8	0.65	
Q8	2,753	4.34	0.91	12.2		0.58
Q9	3,107	4.68	0.67	0.9	0.86	
Q10	3,103	4.70	0.67	1.0	0.85	
Q11	3,015	4.59	0.80	3.8	0.76	
Cronbach's α					0.95	0.87
Percent of variance					68.54	9.10

n = number of responses; S.D. = standard deviation; U/A = unable to assess.

questionnaires for medical colleagues, coworkers and self.¹⁸

The final instrument for patients consisted of 11 items (Table I). Assessors were asked to use a five-point rating scale (1 = strongly disagree to 5 = strongly agree). The instrument for medical colleagues (Table II) and co-workers (Table III) consisted of 19 and 29 items, respectively, with a five-point rating scale (1 = among the worst to 5 = among the best). The self-assessment instrument (Table IV) was identical to the medical colleague instrument except that all items were written in the first person. All questionnaires provided respondents with the option of being able to indicate they were unable to assess the physician on the item.

A private company, Pivotal Research (formerly Customer Information Services), in conjunction with the CPSA, sent invitation letters to all 197 anesthesiologists who were eligible to participate in PAR (i.e., registered as an anesthesiologist and had been in practice for at least three years). Each participating specialist was responsible for completing a self-assessment and identifying the eight medical colleagues and eight coworkers who could answer the questions on the survey. Previous work had established that raters chosen by people being assessed do not provide significantly different evaluations than those selected by a third party.¹¹ Furthermore, studies examining how well the assessor and assessed physician know one another show that 'familiarity' contributes very little

to the variance in ratings.^{7,10} Each anesthesiologist was provided with 30 copies of the patient survey, as previous studies⁶⁻⁸ had shown that surveys from 25 patients would be required to produce reliable data. The physicians were provided with addressed envelopes so that the patients could mail their anonymous responses directly to Pivotal Research. Additional copies of the patient survey would be provided to anesthesiologists if sufficient surveys were not returned.

A number of statistical analyses were undertaken to address the research questions posed. Response rates were used to determine feasibility for each of the respondent groups (Question 1). For each item on each survey, the percentage of 'unable to assess' along with the mean and standard deviation was computed in order to determine the viability of items and the score profiles (Questions 2 and 3, respectively). When the percentage of 'unable to assess' items exceeds 20% on a survey, it may suggest a need to examine the item for revision or deletion. We used exploratory factor analysis to determine which items on each survey belonged together and became a 'factor' (or scale) such as communication (Question 4). This analysis allowed us to identify the factors and the number of factors for each survey and to describe the relative variance accounted for by each factor within the whole instrument. These factors or scales could then be used within the feedback report to guide overall direction for improvement. Individual items within the factor would provide more precise data about specific behav-

iors which the physician might change. As the self data is reported in feedback reports to the physician alongside the medical colleague data, a factor analysis of the self assessment questionnaire was not done. Finally, reliability was assessed (Question 5). Internal consistency reliability was examined using the Cronbach's α coefficient for each of the rater groups and for each of the scales/factors for each rater group. This would tell us whether the instruments have overall stability. This analysis was followed by a generalizability analysis to determine the E_p^2 to ensure there were sufficient numbers of items and raters to provide stable data for each individual anesthesiologist on each instrument. Normally, an E_p^2 of 0.70 suggests data are stable.^{1,5,8} Too low an E_p^2 suggests the need for modifications to the measurement procedure (i.e., more raters or more items).

The study received approval from the Conjoint Health Research Ethics Board of the University of Calgary. The University researchers received anonymous data to conduct the psychometric assessment. The data did not contain any physician characteristics (e.g., location of practice, year or school of graduation) that could be associated with any practitioner.

Results

A total of 197 anesthesiologists were recruited and 186 participated in the study. Of these, 177 anesthesiologists provided a total of 3,135 patient surveys (mean of 17.71, range from 1–32). The majority of anesthesiologists (151 or 76.6%) received 30 questionnaires; the remaining anesthesiologists received anywhere from one to 41 additional copies of the surveys. A total of 182 anesthesiologists provided 1,415 coworker surveys (mean of 7.75, range from 4–8). There were 1,407 medical colleague surveys for 180 anesthesiologists (mean of 7.82, range from 6–8). A total of 182 anesthesiologists returned self-assessment forms. The mean response rates were 95.1% for coworkers, 94.6% for colleagues, 56.2% for patients and 97.8% for the self assessment forms.

The majority of items on the questionnaires could be answered by respondents. As presented in Tables I to IV (Tables III and IV available as Additional Material at www.cja-jc.org) the assessment of 'unable to assess' items showed that two items (of 11) on the patient survey, one on the coworker (of 19), and five (of 29) on the medical colleague survey had unable to respond rates > 15%. The mean ratings for all items on the patient, medical colleague and peer surveys were between 4 and 5.

The factor analysis identified two factors on the patient survey that accounted for 77.6% of the vari-

ance, professionalism and communication. The factor analysis identified two factors (communication and collaboration) on the coworker instrument as accounting for 67.5% of the variance. The medical colleague assessment identified three factors accounted for 74.5% of the variance, clinical performance, communication and professionalism, and continuing professional development.

A Cronbach's α was calculated to determine the internal reliability of the instruments. Patient surveys had an alpha of 0.93, coworker surveys of 0.95, medical colleagues of 0.97 and the self assessment was 0.97. The generalizability coefficients (E_p^2) were 0.65, 0.56, and 0.69 for patient, coworker, and the medical colleague surveys, respectively.

Discussion

In this study we assessed the use of questionnaire based assessments of anesthesia practice. This is the first study which combines feedback from patients, coworkers, and medical colleagues for assessing anesthesiologists.

The PAR program is mandatory and the response rates were high as expected (except for the patient surveys). As such, these rates are consistent with the response rates for other groups of physicians who have been studied.⁷⁻⁹ As noted, the most challenging component was the patient survey. Unlike other groups^{7,8} in which 25 patient surveys were given to the physician to be distributed in the office and virtually all are collected, the mean response rate was < 60% and each anesthesiologist was provided with a minimum of 30 surveys. Despite this, we believe these data show that it is feasible to design a MSF program for anesthesia practice which includes a medical colleague, coworker, patient, and self components. The patient component, however, does need to be regularly reviewed for feasibility.

The majority of the items could be answered by the anesthesiologist's assessors. There were some items which proved difficult for respondents to assess. In some cases the items which were less likely to be answered may have simple explanations which relate to the opportunity to observe the behavior assessed by that item. For example, three of these items on the medical colleague questionnaire were within the area of professional development and participation in quality improvement activities. For anesthesiologists who selected surgeons as respondents, this information may not have been known. On the patient survey, items related to decision making and anesthetic options may not have been recalled subsequently when patients were responding.

TABLE II Descriptive statistics and factor analysis of medical colleague surveys

<i>Items</i>	<i>n</i>	<i>Mean</i>	<i>S.D.</i>	<i>% UA</i>	<i>Factors</i>		
					<i>Clinical performance</i>	<i>Communication and professionalism</i>	<i>Continuing professional development</i>
Q1 Communicates effectively with patients	1,386	4.48	0.65	1.50		0.60	
Q2 Communicates effectively with patients' families	1,194	4.44	0.68	15.10		0.56	
Q3 Communicates effectively with other health care professionals	1,403	4.48	0.71	0.30		0.74	
Q4 Performs technical procedures skillfully within the range of services provided by this physician	1,392	4.55	0.63	1.10	0.67		
Q5 Performs appropriate pre operative assessments	1,370	4.52	0.65	2.60	0.73		
Q6 Formulates an appropriate preoperative plan	1,351	4.53	0.63	4.00	0.73		
Q7 Demonstrates an appropriate level of vigilance intraoperatively	1,360	4.57	0.63	3.30	0.78		
Q8 Uses monitoring devices appropriately during the intraoperative period	1,350	4.54	0.63	4.10	0.77		
Q9 Responds appropriately during clinically significant events	1,370	4.61	0.61	2.60	0.70		
Q10 Selects appropriate interventions (e.g., blood transfusions)	1,343	4.52	0.64	4.50	0.71		
Q11 Communicates effectively with the surgical team intraoperatively	1,343	4.51	0.68	4.50		0.72	
Q12 Seeks help appropriately, perioperatively	1,335	4.51	0.63	5.10	0.62		
Q13 Provides appropriate postoperative care during recovery from anesthetic	1,375	4.52	0.63	3.20	0.71		
Q14 Maintains quality medical records	1,132	4.38	0.69	19.50	0.59		
Q15 Handles transfer of care appropriately	1,305	4.43	0.66	7.20	0.57		
Q16 Maintains confidentiality of patients and their families	1,193	4.51	0.63	15.20	0.62		
Q17 Respects the rights of patients	1,332	4.54	0.64	5.30	0.61		
Q18 Collaborates with medical colleagues	1,382	4.49	0.70	1.80		0.71	
Q19 Is involved with own continuing professional development	1,080	4.48	0.66	23.20			0.75
Q20 Accepts responsibility for own professional actions	1,374	4.53	0.65	2.30		0.060	
Q21 Manages operating room resources efficiently	1,303	4.42	0.70	7.40		0.51	
Q22 Demonstrates punctuality in his/her professional behaviour	1,388	4.43	0.77	1.40		0.60	
Q23 Manages stress effectively	1,227	4.33	0.78	12.80		0.72	
Q24 Recognizes his/her own clinical limitations	1,333	4.46	0.67	5.30		0.61	
Q25 Handles requests for consultation in a timely manner	1,218	4.42	0.69	13.40		0.53	0.53
Q26 Critically appraises sources of medical information	1,119	4.46	0.66	20.50		0.62	
Q27 Facilitates the learning of medical colleagues and coworkers	1,267	4.43	0.69	10.00		0.77	
Q28 Participates in quality improvement programs	1,008	4.41	0.69	28.40			0.78
Q29 Exhibits professional and ethical behaviour toward physician colleagues	1,399	4.55	0.68	0.60		0.75	
Cronbach's α					0.97	0.96	0.91
Percent of variance					67.2	3.8	3.5

n = number of responses; S.D. = standard deviation; U/A = unable to assess.

The range and the mean ratings were similar to that of other groups with most physicians receiving all of their ratings between 4 and 5.⁷⁻⁹ While these scores are high, they are consistent with the range of scores found in most assessments of residents and medical students as well as practicing physicians. Similarly, the self ratings were lower than those provided by medical colleagues, a finding which is consistent with other studies of this nature.^{7,8}

The factor analysis was helpful in identifying the factors (scales) for each of the surveys. Our analysis found two factors for the patient and the coworker surveys and three for the medical colleague survey. These factors of professionalism, communication, collaboration, clinical performance, and continuing professional development allowed us to confirm that our tools had assessed the key domains which the CPSA wanted to examine. It also showed that our scales were similar to those from other PAR instruments,⁶⁻⁹ thus maintaining the integrity of the PAR assessment. The creation of scales through factor analysis also offers the advantage that the scales can be used to guide physicians in global areas (e.g., communication) as well as on an item by item basis (e.g., 'talked with me about anesthetic options'). Participants received feedback about their own performance as well as data for the entire group of anesthesiologists who participated in the study.

The reliability analysis indicates that overall, the instruments are stable. The Cronbach's α was high and the generalizability coefficient, while not as robust as found in our previous work,⁷ was comparable to that found with the American Board of Internal Medicine work.¹⁰ These data suggest that the mix of items and number of raters on the surveys is appropriate and that practitioners can be confident that their feedback is reliable (stable).

Overall, we believe that it is possible to develop high quality MSF instruments for the practice of anesthesia. While this method cannot substitute for direct observation of anesthesiologists^{2,3} or the opportunities to assess clinical performance, teamwork and collaboration provided by simulation methods² or profiling techniques,⁴ these instruments provide unique information about communication skills and professionalism. Moreover, it is a relatively inexpensive method of providing feedback about communication skills, professionalism, collaboration and continuing professional development interest from those who have first hand experience with the clinician, and observe these behaviours directly.

Acknowledgements

Funding for the study was provided by the College of Physicians and Surgeons of Alberta. Data collection was provided by Pivotal Research, Edmonton. We offer special thanks to Robert Burns, John Swiniarski, and Bryan Ward at the CPSA for allowing us to continue to be part of this work. We thank our colleagues Drs. David Archer, Timothy Tang, Rein Paasuke, Harry Patterson, and Brendan Finucane who served on the Anesthesia—PAR working group.

References

- 1 *Royal College of Physicians and Surgeons of Canada*, Information by specialty or subspecialty anesthesia. Available from URL; <http://rcpsc.medical.org/information/index.php>. Accessed July 21, 2005.
- 2 *Fletcher GC, McGeorge P, Flin RH, Glavin RJ, Maran NJ*. The role of non-technical skills in anaesthesia: a review of current literature. *Br J Anaesth* 2002; 88: 418–29.
- 3 *Slagle J, Weinger MB, Dinh MT, Brumer VV, Williams K*. Assessment of the intrarater and interrater reliability of an established task analysis methodology. *Anesthesiology* 2002; 96: 1129–39.
- 4 *St Jacques PJ, Patel N, Higgins MS*. Improving anaesthesiologist performance through profiling and incentives. *J Clin Anesthesiol* 2004; 16: 523–8.
- 5 *Evans R, Elwyn G, Edwards A*. Review of instruments for peer assessment of physicians. *BMJ* 2004; 328: 1240.
- 6 *Lockyer J*. Multisource feedback in the assessment of physician competencies. *J Contin Educ Health Prof* 2003; 23: 4–12.
- 7 *Hall W, Violato C, Lewkonja R, et al*. Assessment of physician performance in Alberta: The physician achievement review. *CMAJ* 1999; 161: 52–7.
- 8 *Violato C, Lockyer J, Fidler H*. Multisource feedback: a method of assessing surgical practice. *BMJ* 2003; 326: 546–8.
- 9 *Lockyer J, Violato C*. An examination of the appropriateness of using a common peer assessment instrument to assess physician skills across specialties. *Acad Med* 2004; 79(10 Suppl.): S5–8.
- 10 *Lipner RS, Blank LL, Leas BF, Fortna GS*. The value of patient and peer ratings in recertification. *Acad Med* 2002; 77(10 Suppl.): S64–6.
- 11 *Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo JP*. Use of peer ratings to evaluate physician performance. *JAMA* 1993; 269: 1655–60.
- 12 *Ramsey PG, Carline JD, Blank LL, Wenrich MD*. Feasibility of hospital-based use of peer ratings to evaluate the performances of practicing physicians. *Acad Med* 1996; 71: 364–70.

- 13 *Le May S, Dupuis G, Harel F, Taillefer MC, Dube S, Hardy JF.* Clinimetric scale to measure surgeons' satisfaction with anesthesia services. *Can J Anesth* 2000; 47: 398–405.
- 14 *Carnie J.* Patient feedback on the anaesthetist's performance during the pre-operative visit. *Anaesthesia* 2002; 57: 697–701.
- 15 *Thoms GM, McHugh GA, Lack JA.* What information do anaesthetists provide for patients. *Br J Anesth* 2002; 89: 917–9.
- 16 *Dexter F, Aker J, Wright WA.* Development of a measure of patient satisfaction with monitored anesthesia care. The Iowa Satisfaction with Anesthesia Scale. *Anesthesiology* 1997; 87: 865–73.
- 17 *Fidler H, Lockyer JM, Toews J, Violato C.* Changing physicians' practices: the effect of individual feedback. *Acad Med* 1999; 74: 702–14.
- 18 *College of Physicians and Surgeons of Alberta,* Physician Achievement Program. Available from URL; <http://www.par-program.org/> (accessed July 21, 2005).