

An Examination of the Appropriateness of Using a Common Peer Assessment Instrument to Assess Physician Skills across Specialties

JOCELYN M. LOCKYER and CLAUDIO VIOLATO

ABSTRACT

Problem Statement. To determine whether a common peer assessment instrument can assess competencies across internal medicine, pediatrics, and psychiatry specialties.

Method. A common 36 item peer survey assessed psychiatry ($n = 101$), pediatrics ($n = 100$), and internal medicine ($n = 103$) specialists. Cronbach's alpha and generalizability analysis were used to assess reliability and factor analysis to address validity.

Results. A total of 2,306 (94.8% response rate) surveys were analyzed. The Cronbach's alpha coefficient was .98. The generalizability coefficient (mean of 7.6 raters) produced an $Ep^2 = .83$. Four factors emerged with a similar pattern of relative importance for pediatricians and internal medicine specialists whose first factor was patient management. Communication was the first factor for psychiatrists.

Conclusions. Reliability and generalizability coefficient data suggest that using the instrument across specialties is appropriate, and differences in factors confirm the instrument's ability to discriminate for specialty differences providing evidence of validity.

The value of peer feedback to assess and inform physicians about a broad range of competencies has been well established.¹⁻⁷ Interest in tools to assess peer performance has gained momentum recently as part of worldwide interest in improving patient safety^{2,8} and the need to develop tools to assess ACGME competencies in the United States.^{2,6,8} Licensing authorities,^{1,2,7,8} professional organizations,⁶ and health care facilities⁴ have adopted peer feedback as part of quality improvement^{1,2} and maintenance of certification programs.⁶ Peer feedback has been shown to be useful in assessing communication skills, interpersonal skills, collegiality, medical expertise, and the ability to continually learn and improve practice patterns,^{2,3,8} competencies that medical organizations and the public believe need attention.^{9,10}

Studies show that questionnaire-based assessment by peers can be reliable, valid, and demonstrate good psychometric properties.^{1,2,4-7} They are feasible, with most structured so that ten or fewer peers assess a physician, a number that the generalizability coefficient data suggest produces an acceptable Ep^2 (i.e., $> .7$).^{1,2,5,8} From a learning and change perspective, physicians do make changes in their practices in response to this feedback.^{2,6,8,11}

The development and testing of reliable, valid, and feasible instruments is resource intensive. Careful attention has to be paid to instrument psychometrics as well as to the collegial and democratic processes used to develop, disseminate, and communicate the program with physicians who will be assessed and assess. Clear objectives for the program; procedures for collecting, handling, and reporting data; as well as clarity about the repercussions that might result from suboptimal performance assessment must be clearly articulated.⁸ Peer review provokes anxiety, and physician perception of the value of the data may be undermined if physicians do not

believe their raters capable of the assessments being made.^{5,8} Nonetheless, in most settings, it is impractical to develop specialty, subspecialty, or clinic-specific instruments that are psychometrically sound and feasible. It thus becomes important to assess whether a single generic peer instrument can be used across specialties to assess competencies.

The main purpose of the present study was to determine whether a common 36-item peer assessment instrument provides a reliable and valid assessment of competencies across the three specialties of internal medicine (IM), pediatrics, and psychiatry. The study addressed the following questions for each of the three specialty groups: (1) What is the reliability of the instrument and its scales/factors? (2) What are the generalizability coefficients (Ep^2)? (3) Which factors and how many factors could be identified? (4) What is the relative variance accounted for by each factor as well as their coherence and theoretical interpretability? (5) How do the factor structures compare across specialty groups? (6) Is the instrument sensitive to the differences and similarities inherent in the practice of each specialty?

Method

The peer instrument studied is one component of the College of Physicians and Surgeons of Alberta, Physician Achievement Program (CPSA-PAR) begun in 1996^{1,2,12} in which medical colleagues, co-workers, and patients provide assessments of physicians in practice. The goal 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.¹ Instruments had been developed and tested for family physicians^{1,12} and surgeons^{2,12} but not for pediatric, IM, or psychiatry specialists.

A working group was recruited to develop a set of instruments for these physicians that would be similar to previously developed instruments that had been found to be reliable and valid.^{1,2,8} The instruments were to contain some new items to reflect contemporary ideas about physician competence^{3,9,10} and include the context of a referral practice, professional development, and the quality improvement/safety concerns of those specialties. The preference was to use the same instrument for all three specialty groups. After the working group determined the items for inclusion on the instruments, copies of the instruments were sent to all specialists in IM, pediatrics, and psychiatry for review and feedback. The working group made adjustments to the instruments based on this feedback. The final instrument (Table 1) consisted of 36 items. Raters were asked to use a five-point rating scale (1 = among the worst to 5 = among the best) to assess the physician. The raters had the option indicating they were unable to assess their peer on items.

Data from a sample of 304 specialty physicians ($n = 101$ psychiatrists, $n = 100$ pediatricians, $n = 103$ IM specialists), recruited by the CPSA-PAR Program, were used to assess the medical colleague instrument. All specialists in the study sample had been in practice a minimum of five years. Each participating specialist was responsible for identifying eight medical colleagues who could answer the questions on the survey on their behalf. Recognizing the diversity of practices, the colleagues could be any combination of peer, referring, or referral physicians.

TABLE 1. Factor Analysis of 36-Item Peer Assessment Instrument Used to Assess Psychiatrists, Internal Medicine, and Psychiatry Specialists

Item	Internal Medicine				Pediatrics				Psychiatry			
	PM*	CA*	PD*	Com*	PM*	CA*	PD*	Com*	Com*	PM*	CA*	PD*
Handles requests for consultation in a timely manner	.756				.681					.711		
Recognizes his/her limitations	.740				.747				.512	.483		
Advises referring physician if referral request is outside the scope	.721				.683					.585	.414	
Manages health care resources efficiently	.698				.587					.489		
Provides a clear understanding about who is responsible for the continuing care of patients	.691				.735					.576	.453	
Coordinates care effectively for patients with other health care professionals and physicians	.664				.570		.406		.479	.447		
Gives priority to urgent requests	.663					.553	.411			.735		
Handles transfer of care appropriately	.661		.401		.673					.584	.495	
Maintains confidentiality of patients and their families	.635				.546			.401	.432	.479		
Manages own stress effectively	.633		.420		.668					.484		
Assumes appropriate responsibility for patients	.632		.421		.487					.613		
Respects the rights of patients	.630		.435		.484			.440	.614	.420		
Participates in a system of call to provide care for his/her own patients when unavailable	.623		.403		.592					.698		
Makes appropriate use of community resources for psychosocial aspects of care	.591			.496	.587				.445	.471		
Shows compassion for patients and their families	.585			.481			.497	.510	.745			
Provides timely information to referring physicians about mutual patients	.579				.724					.578	.402	
Exhibits professional and ethical behavior towards physician colleagues	.570	.411			.630		.469		.533	.551		
Recognizes psychosocial aspects of illness	.563			.507	.435			.401	.588			
Accepts responsibility for own professional actions	.561		.468		.544		.465		.448	.501		.428
Collaborates with physician colleagues	.537		.419	.426	.520		.533		.528	.456		.403
Maintains quality records	.511	.409		.496	.586	.404					.622	
Handles emergency situations effectively	.505	.463	.429		.431	.608				.588		
Participates effectively as a member of the health care team	.498		.479		.466		.628		.541	.429		.467
Selects diagnostic tests appropriately		.751			.407	.618					.664	
Makes the correct diagnosis following consultation		.745				.740			.470		.628	
Critically assesses diagnostic information		.717				.648					.700	
Selects appropriate treatments within the range of services provided by this physician, he/she performs technical procedures skillfully		.716				.671			.458		.618	
		.616				.641					.635	
Is involved with professional development			.699				.733					.732
Contributes to quality-improvement programs and practice guidelines	.404		.665				.742					.756
Facilitates the learning of medical colleagues and co-workers			.662				.777					.745

(Continues)

TABLE 1. (Continued)

Item	Internal Medicine				Pediatrics				Psychiatry			
	PM*	CA*	PD*	Com*	PM*	CA*	PD*	Com*	Com*	PM*	CA*	PD*
Critically evaluates the medical literature to optimize clinical decision making		.466	.645				.680					.658
Manages patients with complex problems			.639			.618	.423		.411		.480	
If a member of my own family needed care, I would rate this physician		.446	.521			.490			.663			
Is able to verbally communicate effectively with other health care professionals				.805				.871	.756			
Communicates effectively with patients				.759				.839	.764			
Writes legibly		.463		.636				.654	.596		.463	
Communicates effectively with other health care professionals				.630	.558			.445	.579			
Cronbach's alpha	.94	.92	.96	.91	.97	.93	.92	.94	.93	.95	.91	.93
% of variance	63.2	4.3	3.2	2.7	56.0	4.3	3.8	3.5	56.2	3.9	3.5	3.2

*PM = patient management; CA = clinical assessment; PD = professional development; Com = communication.

To assess whether the peer instrument could provide a valid and reliable assessment of competencies across specialties, a number of statistical analyses were performed using the data for each specialty group. Internal consistency reliability was assessed using the Cronbach's alpha coefficient for each of the physician groups and for each of the scales/factors for each physician group. A generalizability analysis was done to determine the generalizability coefficient (Ep^2). The latter analysis is required to ensure there are sufficient numbers of items and raters to reach at least ($Ep^2 = .70$).^{1,5,8} Too low an Ep^2 suggests the need for modifications to the measurement procedure, as this assessment can determine the sources of measurement error as well as the number of items and observers required to obtain a desired level of generalizability.

Factor analyses were done to identify the underlying common set of variables for each specialty. Exploratory factor analysis allows one to determine which items belong together (i.e., are a "factor"), the patterns of order of items within the factor, and the commonalities and differences for each of the three specialty groups. This type of analysis is used to investigate common but unobserved sources of influence in a collection of variables. Its empirical basis is the observation that variables from a carefully chosen domain are often intercorrelated. Thus, it is natural to hypothesize that the variables all reflect a more fundamental influence that contributes to individual differences on each measure.^{13,14} Factor analysis, in this study, was used to decompose the variability of items into two parts: one part attributable to a factor and shared with other items, and a second part that is specific to that item but unrelated to other factors.

Using individual-physician data as the unit of analysis, the 36 items were intercorrelated using Pearson product-moment correlations. The correlation matrix was then decomposed into principle components and these were subsequently rotated to the normalized varimax criterion. This principal component extraction with varimax rotation was employed to determine the factor structure of the instrument for each specialty and the appropriateness of the items for assessing those factors. Items were considered to be part of a factor if their primary loading was on that factor. The number of factors to be extracted was based partly on the Kaiser rule (i.e., eigenvalues > 1.0) and results from our previous research.^{1,2} With three specialty groups, a comparison of factors by specialty group would allow us to identify group differences as well as similarities, even though the same instrument was used. Thus factor analysis

would allow us to identify the factors and numbers of factors for each specialty group, describe the relative variance accounted for by each factor, their coherence and theoretical interpretability, compare factors across specialty groups and determine whether the instrument was sensitive to differences and similarities in the practice of the different groups.

The study received approval from the Conjoint Health Research Ethics Board of the University of Calgary.

Results

A total of 2,306 peer surveys out of a possible 2,432 (94.8%) were completed for the 304 participants in the study. The 101 psychiatrists provided a mean of 7.56 surveys (range of 5–8) for a total of 764 (94.6%) surveys. The 100 pediatricians provided a mean of 7.64 surveys (minimum of 5 and maximum of 8) for a total of 764 (95.5%) surveys. The 103 IM specialists provided a mean of 7.64 surveys (minimum of 6 and maximum of 8) for a total of 778 (94.4%) surveys.

The generalizability coefficient (Ep^2) was calculated using the entire sample of physicians as well as for each specialty group. The two facets used in this analysis were raters and items on the instrument. The mean number of peer assessors across all physicians was 7.6 producing an Ep^2 of .83 (for pediatricians $Ep^2 = .78$; for psychiatrists $Ep^2 = .81$; and for IM specialists $Ep^2 = .82$).

The factor analysis for psychiatrists showed that four eigenvalues were greater than one (21.4, 1.5, 1.3, 1.2). Nearly 70% of the variance (66.86%) was "accounted for" by this solution. The varimax rotation converged in 12 iterations. Factor 1 (communication skills) accounted for 56.2% of the total variance, Factor 2 (patient management) for 3.9%, Factor 3 (clinical assessment) for 3.5%, and Factor 4 (professional development) for 3.2%. For pediatricians, four eigenvalues were greater than one (21.3, 1.6, 1.4, 1.3) and accounted for 67.6% of the variance. The varimax rotation converged in eight iterations. Factor 1 (patient management) accounted for 56.0% of the total variance, Factor 2 (clinical assessment) for 4.3%, Factor 3 (professional development) for 3.8%, and Factor 4 (communication) for 3.5%. For IM, four eigenvalues were greater than one (24.0, 1.6, 1.2, and 1.0). This accounted for 73.4% of the variance. Factor 1 (patient management) accounted for 63.2% of the total variance, Factor 2 (clinical assessment) for 4.3%,

Factor 3 (professional development) for 3.2%, and Factor 4 (communication) for 2.7%. See Table 1.

A Cronbach's alpha coefficient was completed to assess reliability for each specialty and each factor. For psychiatrists, the Cronbach's alpha coefficient was .98 with an average standard error of measurement (SEM) of .17. For pediatricians, the Cronbach's alpha coefficient was .98 and the SEM was .18. For IM specialists, the Cronbach's alpha coefficient was .99 and the SEM was .09. For the factors for each specialty, the Cronbach's alpha was >.90 (see Table 1).

The items that aligned with each of the factors for all specialty groups were similar. Patient management encompassed those aspects of specialist practice in which the physician provides consulting expertise for the patient's care, manages resources, and overall care coordination. Clinical assessment encompassed items like selecting and using diagnostic information in treatment choice. The professional development factor included involvement with professional development, contributing to quality improvement programs, and facilitating learning for others. The communication factor included verbal communication with medical colleagues, other health professionals, and patients.

Discussion

The response rate is extremely high and consistently high for all three groups. This is due in large part to the fact that the CPSA-PAR program in Alberta is mandatory. The Cronbach's analyses for all groups provide evidence for high internal consistency reliability. The average standard error of measurement ranged from .09 to .18, showing that the instruments had good distributional and psychometric properties for item discrimination (i.e., discerning between different physicians). These results were consistent with our previous studies of family physicians and surgeons.^{1,2} We achieved a high $Ep^2 > .80$ with between seven and eight raters per physician. This confirms the appropriateness of using fewer raters but more items on the instrument and contrasts with the instruments used by the American Board of Internal Medicine, which uses fewer items but requires more raters.⁴⁻⁶ In Alberta, the CPSA's goal in assessing physicians was to provide quality improvement data on a full range of competencies.¹ There are advantages in a province with limited numbers of total physicians to reduce the number of times each physician may be called upon to assess his or her colleagues while maximizing the practice information (i.e., items of data) provided to each physician who is assessed.

The factor analyses revealed the same four factors for all three physician-specialty groups: patient management, clinical assessment, professional development, and communication. While these groupings are similar for all three groups, they reveal some important differences. The pattern of importance of the factors (amount of variance accounted for and cohesiveness) was the same for pediatricians and IM specialists but different for psychiatrists. For IM specialists and pediatricians, patient management was the most important factor while communication was the most important factor for psychiatrists. This was followed by patient management skills for psychiatrists. These differences likely reflect the differences in practices among the three physician groups. Effective communication is generally considered to be the foundation of psychiatry¹⁵⁻¹⁷ and its "daily business." Psychiatry values communication skills above all else as these are central to diagnosis and assessment as well as therapy.¹⁷ A psychiatrists' ability to communicate will determine his/her effectiveness in eliciting information and establishing and carrying outpatient care interventions related to the patient's mental health problems over relatively long periods. Conversely, the nature of Canadian IM and pediatric referral practice is

often based on managing urgent requests from physician colleagues, using on his or her diagnostic skills and information to establish a protocol for the patient, stabilize the patient, and return the patient to the family physician for ongoing care. These findings related to similarities and differences in factors provide both convergent and divergent evidence for the validity of peer assessments in the present study.

In conclusion, our results indicate that the instrument is appropriate for use across the three specialties. The tools are sensitive to the differences inherent in the practice of psychiatry and emphasize the key components of the specialized consultative nature of IM, pediatrics, and psychiatry. The psychometric quality of the instruments is high. Furthermore, while these tools were developed for use in a Canadian context, they may be useful in other jurisdictions. The family physician instruments have been used in a pilot project in Nova Scotia.⁷ Medical organizations in Germany, New Zealand, Malaysia, France, and California have all indicated their desire or intention to use the instruments (personal communication) developed by the CPSA-PAR program.¹²

Funding for the study was provided by the College of Physicians and Surgeons of Alberta. Data collection was provided by Customer Information Services, Edmonton. Special thanks to Robert Burns, John Swinarski, and Bryan Ward at the CPSA for allowing us to continue to be part of this work. At the University of Calgary, we thank our colleagues Herta Fidler, John Toews, Ray Lewkonja, and Keith Brownell for their ongoing interest and review of documents from our multisource (360-degree) studies.

Correspondence: Jocelyn Lockyer, University of Calgary, 3330 Hospital Drive NW, Calgary, AB. T2N 4N1; e-mail: {lockyer@ucalgary.ca.}

References

1. Hall W, Violato C, Lewkonja R, et al. Assessment of physician performance in Alberta: the Physician Achievement Review. *CMAJ*. 1999;161:52-7.
2. Violato C, Lockyer J, Fidler H. Multi source feedback: a method of assessing surgical practice. *BMJ*. 2003;346-8.
3. Norcini JJ. Peer assessment of competence. *Med Educ*. 2003;37:539-43.
4. 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.
5. 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.
6. Lipner RS, Blank LL, Leas BF, Fortna GS. The value of patient and peer ratings in recertification. *Acad Med*. 2002;77(suppl 10):S64-6.
7. Sargeant JM, Mann KV, Ferrier SN, Langille DB, Muirhead PD, Sinclair DE. Responses of rural family physicians and their colleagues and coworker raters to a multi-source feedback process: a pilot study. *Acad Med*. 2003;78(suppl 10):S42-4.
8. Lockyer J. Multisource feedback in the assessment of physician competencies. *J Contin Educ Health Prof*. 2003;23:23:4-12.
9. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA*. 2002;287:226-35.
10. Levine AM. Medical professionalism in the new millennium: a physician charter. *Ann Intern Med*. 2002;136:243-6.
11. Fidler H, Toews J, Lockyer J, Violato C. Changing physicians' practices: The effect of individual feedback. *Acad Med*. 1999;74:702-14.
12. College of Physicians and Surgeons of Alberta. Physician Achievement Program (<http://www.par-program.org/>). Accessed 20 January 2004.
13. Russell DW. In search of underlying dimensions. The use (and abuse) of factor analysis in *Personality and Social Psychology Bulletin*. *Pers Soc Psych Bull* 2002;28:1629-46.
14. Cudeck R. Exploratory factor analysis. In: Tinsley HEA, Brown SD (eds), *Handbook of Multivariate Statistics and Mathematical Modeling*. San Diego: Academic Press, 2000:95-124.
15. Schreiber SC, Kramer TA, Adamowski SE. The implications of core competencies for psychiatric education and practice in the US. *Can J Psychiatry*. 2003;48:215-21.
16. Tuhan J. Mastering CanMEDS roles in psychiatric residency: a resident's perspective. *Can J Psychiatry*. 2003;222-4.
17. Martin L, Saperson K, Maddington B. Residency training: challenges and opportunities in preparing trainees for the 21st century. *Can J Psychiatry*. 2003;48:225-30.