Moderator: Barbara Barnes, MD
Discussant: Larry Gruppen, PhD

What Multisource Feedback Factors Influence Physician Self-Assessments? A Five-Year Longitudinal Study

Jocelyn M. Lockyer, Claudio Violato, and Herta M. Fidler

Abstract

Background
Multisource feedback, in which medical colleagues, patients, coworkers, and the physician involved provide data, is a tool to inform physician practice. Its impact on physicians' self-assessment through two iterations is unknown.

Method
Data from 250 family physicians in Alberta who participated in two iterations, five years apart—1999 and 2006—allowed the authors to determine the change in self-assessment scores, using a *t* test. A multiple regression was used to account for the variance in the scores from the second self-assessment by the data from the multisource feedback and sociodemographics from the first iteration.

Results
Physicians rated themselves higher in the second iteration. The linear regression model accounted for 27.4% of the variance in the ratings at the second iteration and incorporated data from the self-assessment.

Conclusions
Physician self-assessment seems driven by stable perceptions that physicians hold about themselves and that may be slow to change.

Acad Med. 2007;82(10 Suppl):S77–S80.

The accuracy of physician self-assessment skill has been questioned in a recent, systematic review of medical education literature.[1] This study showed that 13 out of 20 comparisons of self- and external comparisons showed a small, none, or an inverse relationship. Only seven demonstrated positive associations. Multisource feedback studies in which physicians complete a self-assessment questionnaire, identical in content to that provided by medical colleagues, provides additional evidence that physicians self-assessment scores are divergent from those of their medical colleagues.[2] Accordingly, it seems that physicians in general are not very accurate in assessing themselves in comparison with external sources such as performance or peer assessments.

Self-assessment is an important skill. The belief that physicians should possess the skill to accurately assess themselves and access the resources necessary for improvement underpins the concept of professional self-regulation.[3] Most continuing professional development for physicians is predicated on the assumption that physicians are self-directed learners who can accurately determine their own learning needs, set goals, engage in appropriate learning activities, and assess the outcomes of their learning.

It has been suggested, however, that we need to move beyond "guess your grade" studies of physician self-assessment skill. Attention needs to focus on the pedagogical/practice-based impact of engaging in formal self-assessment activities, physician feedback-seeking behavior, sources used by physicians to assess themselves, and the impact of the feedback on practice.[4] Practicing physicians have few regular sources of feedback. Multisource feedback has been suggested as a method of enhancing physician self-assessment skills[1] and may be particularly useful when interpersonal skills, communication skills, or professionalism needs to be evaluated.[2,5] It has been shown that physicians will use the feedback to inform practice change.[5,6] Some physicians fail, however, to use the feedback.[5,6] It seems that the nature of the feedback, emotional reactions to the feedback, the feedback procedures, self-efficacy, and the culture and environment all affect how the feedback will be used.[6] Notwithstanding considerable research that has been done in the area of self-assessment,[1–4,7,8] our understanding of self-assessment is limited. Furthermore, the impact of participation on self-assessment ratings is not known, making it difficult to identify interventions that may provide compelling evidence for change. Accordingly, further research on factors that influence physician self-assessment is required.

Physicians in Alberta have participated in multisource feedback since 1996.[9,10] Questionnaire-based data are provided by eight medical colleagues, 25 patients, and eight coworkers. The physician completes a self-assessment. Each physician is required to participate every five years. Between January 1999 and November 2000 and May 2005 and May 2006, a total of 488 physicians provided data on two occasions. The purposes of the present longitudinal study were to determine (1) whether there was a change in self-assessment scores on the two iterations, and (2) whether the data from the self-, medical colleague, patient, and coworker assessments and sociodemographic data in the first iteration help to explain the self-assessment scores in the second iteration.

Correspondence: Jocelyn M. Lockyer, PhD, Continuing Medical Education and Professional Development, 3330 Hospital Dr NW, Calgary, AB, Canada, T2N 4N1; e-mail: (lockyer@ucalgary.ca).

Method

Pivotal Research Ltd, a company which administers the College of Physicians and Surgeons of Alberta Physician Achievement Review program (PAR), provided an anonymous dataset for this study. We requested a randomly selected sample of 250 (of 488) family physicians/general practitioners who had participated in the program on two occasions. Data for this study consisted of four datasets: self, patients, coworkers, and medical colleagues for iteration one and the self-data for iteration two, five years later. Sociodemographic data provided for each physician included gender (male/female), year of graduation from medical school, school of graduation (Canadian/international), location of practice (urban, regional, rural), and certification status (yes/no).

The medical colleague and self-assessment instruments had 31 items, with the self-assessment instrument written in first person. The coworker (e.g., nurses, pharmacists) assessment had 17 items. The patient assessment had 40 items. Copies of the questionnaires can be found at (www.par.program.org). All assessments used a five-point scale, with the option of indicating that an item was not applicable. Sum mean ratings were calculated for each of the instruments by first aggregating the data and then computing a sum rating by adding the aggregate data for each physician.

To determine whether the self-assessment scores changed between iterations (time one and time two), we used a paired *t* test to compare the scores. We followed that with a Cohen's *d* effect size calculation.

To determine whether the data from the self, medical colleague, patient, and coworker assessments and sociodemographic data in the first iteration explain the self-assessment scores in the second iteration, we intercorrelated all the factors in the first iteration using a Pearson product–moment correlation coefficient to assess for multicollinearity, and we followed this up with a forward linear regression. The factor analysis (FA) was carried out using time one medical colleague, patient, and coworker data. FA was done to determine which of the items in each instrument belonged together (i.e., were a factor). Using individual physician data as the unit of analysis, the items for each of the instruments were intercorrelated using Pearson product–moment correlations. The correlation matrix was then decomposed into principal components, and these were subsequently rotated to the normalized varimax criterion. 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 on the Kaiser rule (i.e., eigenvalues >1.0). Because the medical colleague and self-questionnaires had the same items, we used the medical colleague factor structure for the first iteration of the self-assessment. Finally, we completed the forward multiple-regression analysis, using the sum of the mean score for the second iteration of the self-assessment as the dependent variable. The independent variables included the data from the factors of the first-iteration surveys and the sociodemographic data. Collinearity concerns were addressed using a collinearity diagnostic process.

The conjoint health research ethics board at the University of Calgary approved the study.

Results

The 250 physicians in this sample included 81 (32.4%) females and 169 (67.6%) males. Their median year of graduation was 1980. The majority of physicians (n = 183 [73.2%]) were Canadian graduates, whereas others (n = 67 [26.8%]) were international medical graduates. The physicians described their place of practice as rural (23.6% [n = 59]), regional (10.8% [n = 27]), and urban (65.6% [n = 164]). The sum mean ratings for each of the instruments in the first iteration were 118.70/155 (SD = 17.27) for self, for medical colleague 133.60/155 (SD = 8.70), for coworker 75.60/85 (SD = 5.09), and for patient 180.89/200 (SD = 6.77).

The paired-sample *t* test to compare the sum of the mean scores of the two iterations of the self-assessments indicates a significant difference ($t_{249} = -7.204$; $P < .05$). The sum mean rating the physicians gave themselves in time one (mean = 3.86; SD = 0.51) was statistically significantly lower than their rating of themselves in time two (mean = 4.10; SD = 0.55) and showed a moderate effect size (Cohen's $d = 0.46$).

The intercorrelation matrix showed that the factors on different instruments had low correlations (i.e., $r < 0.16$) and were not significant. Those data, as well as the diagnostics for collinearity, indicate that multicollinearity of independent variables was not a problem.

The FA resulted in 18 factors, three (professionalism and communication, clinical competency, and psychosocial) for the time one medical colleague data, which accounted for 71% of the variance of the scale. The time one patient data encompassed four factors (professionalism and communication, office personnel, access to physicians, and physical office) accounting for 74.8% of the variance. Two factors for the time one coworker data (professionalism and communication) accounted for 67% of the variance. Table 1 provides examples of items for each of the factors, along with the Cronbach's alpha for the factor and the number of items in the factor.

The optimal linear regression model identified two independent variables (self—professionalism and communication, $\beta = 0.379$; P < .000; and self—psychosocial, $\beta = 0.173$; $P < .041$), which accounted for 27.4% of the variance (multiple $R = 0.524$) in time two self-assessment ($F_{2,247} = 46.63$; $P < .001$). None of the sociodemographic variables, the coworker, the medical colleague, or the patient data made significant contributions to the variance in time two self-assessment scores.

Discussion

The present longitudinal study was based on a group of 250 physicians who had completed two iterations of the PAR program, five years apart. Our sample of 250 physicians slightly underrepresented the females in the province (32.4% female in our study versus 39.9% in the province) but was similar for the rural/urban mix (23.6% in our study versus 24.4% in the population as a whole practiced in rural communities).

The main findings are that (1) physicians rated themselves higher in the second iteration than the first iteration, and (2) we were able to account for 27.4% of the variance in the ratings at time two with independent variables which drew from the previous self-assessment data. Physicians seemed not to incorporate

Table 1
Summary of Factor Analyses for Physician Achievement Review Instruments

Questionnaire and factor	Examples of items	Number of items	Cronbach's alpha
Medical colleague—professionalism and communication	Shows compassion for patients and their families; respects the rights of patients; communicates effectively with patients.	16	0.965
Medical colleague—clinical competence	Critically assesses diagnostic information; within the range of services provided by this physician, he/she performs technical procedures skillfully	10	0.937
Medical colleague—psychosocial	Makes appropriate use of community resources for psychosocial aspects of care; manages patients with complex psychosocial problems; makes appropriate referrals for psychosocial aspects of illness	5	0.901
Self—professionalism and communication	As above for medical colleague	16	0.946
Self—clinical competency	As above for medical colleague	10	0.914
Self—psychosocial	As above for medical colleague	5	0.855
Coworker—professionalism	Respects the professional knowledge and skills of coworkers; collaborates well with coworkers	15	0.957
Coworker—communication	Writes prescriptions clearly; written communication with other health professionals is effective	2	0.743
Patient—professionalism and communication	Talks with me about treatment plans; answers my questions well; your doctor explained your illness or injury thoroughly	24	0.983
Patient—office personnel	Staff behaves in a professional manner; staff is pleasant and helpful	6	0.976
Patient—office management	I am able to reach a doctor by telephone after office hours; in urgent cases, a doctor is available by phone	5	0.857
Patient—physician office	Office is easy to get into; office provides adequate privacy	5	0.894

data from medical colleagues, coworkers, or patients into their self-assessment.

The scores in the second iteration were higher than the initial self-assessment scores and were in greater alignment with the higher scores provided by medical colleagues, patients, and coworkers in the first iteration; the linear regression analysis suggests that feedback from medical colleagues, coworkers, or patients did not influence the self-assessment ratings in the second iteration. The major determinants of self-assessment were two of the three factors from the self-assessment questionnaire—namely, those related to professionalism and communication and clinical competency had the highest beta weights. It is likely that physician self-assessment is driven by fairly stable perceptions that physicians hold about themselves. As Eva and Regehr[4] have suggested, self-assessment is a situationally bounded cognitive process that is context specific and dependent on expertise. It is based on perceptions of self-efficacy (belief in one's capability to manage prospective situations), cognitive and meta-cognitive theory (how judgments about one's own knowledge are made), and social cognition (what we know about ourselves resides outside of conscious awareness, and this guides behavior, motivations, and feelings). This finding is consistent with other studies which have shown that self-assessment ratings change very little, even with mentoring and coaching.[8,11] Furthermore, as Sargeant et al[6] have found, physicians' acceptance of multisource feedback depends on many factors, including their emotional reactions to the feedback, the congruence between the feedback and their personal beliefs about themselves, and the nature and characteristics of the feedback itself.

There are limitations to the study. There was a large gap in time (five years) between iterations. It is hoped that data from multisource feedback used with postgraduate trainees who are assessed annually using this approach[12] can be used to further inform the role that first-iteration data have on subsequent self-assessment ratings. Multisource feedback in this particular setting is accompanied by very little formal opportunity to discuss results, set goals, and receive feedback within a shorter time frame. It is possible that a formal mentoring or coaching system in conjunction with the reports would alter self-assessment. Last, it would also be helpful in this type of work to have objective measures of how well these physicians are actually doing in practice (e.g., from clinical outcome and direct observational data), so that physicians had data that complemented the multisource feedback information.

To the best of our knowledge, this is the first longitudinal study of multisource feedback of physicians. The present results suggest that physicians may be slow to incorporate feedback over time. It is possible that with more frequent feedback, mentoring, or other data, physicians will incorporate more data into their self-assessment, but this will require additional testing.

Acknowledgments

The authors acknowledge the ongoing support provided by the College of Physicians and Surgeons of Alberta, namely, John Swiniarski, Bryan Ward, and Trevor Theman; as well as Pivotal Research Inc, namely, Kim Thornton.

References

1 Davis D, Mazmanian P, Fordis M, Harrison RV, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a

systematic review. JAMA. 2006;296:1094–1102.

2 Violato C, Lockyer J. Self and peer assessment of pediatricians, psychiatrists and medicine specialists: implications for self-directed learning. Adv Health Sci Educ Theory Pract. 2006;11:235–244.

3 Ward M, MacRae H, Schlachta C, et al. Resident self-assessment of operative performance. Am J Surg. 2003;185:521–524.

4 Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. Acad Med. 2005;80(10 suppl):S46–S54.

5 Fidler H, Lockyer J, Toews J, Violato C. Changing physicians' practices: the effect of individual feedback. Acad Med. 1999;74:702–714.

6 Sargeant J, Mann K, Sinclair D, van der Vleuten C, Metsemakers J. Understanding the influence of emotions and reflections upon multi-source feedback acceptance and use. Adv Health Sci Educ Theory Pract. 2006; November 8 [Epub ahead of print].

7 Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol. 1999;77:1121–1134.

8 Hodges B, Regehr G, Martin D. Difficulties in recognizing one's own incompetence: Novice physicians who are unskilled and unaware of it. Acad Med. 2001;76(10 suppl):S87–S89.

9 Violato C, Marini A, Toews J, Lockyer J, Fidler H. Feasibility and psychometric properties of using peers, consulting physicians, co-workers, and patients to assess physicians. Acad Med. 1997;72(10 suppl):S82–S84.

10 Hall W, Violato C, Lewkonia R, et al. Assessment of physician performance in Alberta: the Physician Achievement Review. CMAJ. 1999;161:52–57.

11 Eva KW, Cunnington JP, Reiter HI, Keane DR, Norman GR. How can I know what I don't know? Poor self assessment in a well-defined domain. Adv Health Sci Educ Theory Pract. 2004;9:211–224.

12 Archer J, Norcini J, Southgate L, Heard S, Davies H. mini-PAT (Peer Assessment Tool): a valid component of a national assessment programme in the UK? Adv Heal Sci Educ Theory Pract 2006; November 8 [Epub ahead of print].