The Reliability, Validity, and Feasibility of Multisource Feedback Physician Assessment: A Systematic Review

Tyrone Donnon, PhD, Ahmed Al Ansari, MBBCh, MRCSI, PhD, Samah Al Alawi, MD, and Claudio Violato, PhD

Abstract

Purpose
The use of multisource feedback (MSF) or 360-degree evaluation has become a recognized method of assessing physician performance in practice. The purpose of the present systematic review was to investigate the reliability, generalizability, validity, and feasibility of MSF for the assessment of physicians.

Method
The authors searched the EMBASE, PsycINFO, MEDLINE, PubMed, and CINAHL databases for peer-reviewed, English-language articles published from 1975 to January, 2013. Studies were included if they met the following inclusion criteria: used one or more MSF instruments to assess physician performance in practice; reported psychometric evidence of the instrument(s) in the form of reliability, generalizability coefficients, and construct or criterion-related validity; and provided information regarding the administration or feasibility of the process in collecting the feedback data.

Results
Of the 96 full-text articles assessed for eligibility, 43 articles were included. The use of MSF has been shown to be an effective method for providing feedback to physicians from a multitude of specialties about their clinical and nonclinical (i.e., professionalism, communication, interpersonal relationship, management) performance. In general, assessment of physician performance was based on the completion of the MSF instruments by 8 medical colleagues, 8 coworkers, and 25 patients to achieve adequate reliability and generalizability coefficients of $\alpha \geq 0.90$ and $Ep^2 \geq 0.80$, respectively.

Conclusions
The use of MSF employing medical colleagues, coworkers, and patients as a method to assess physicians in practice has been shown to have high reliability, validity, and feasibility.

Assessment and maintenance of physician competence is greatly important to physician organizations. This is particularly true given growing concerns for patient safety[1] and the understanding that professional roles and responsibilities, including interpersonal skills and professionalism, should be integrated into physicians' clinical practice.[2] Thus, the view of competence has changed from a focus on the ability to conduct specific medical procedures to a more comprehensive framework for the assessment of physician performance.[3] Multisource feedback (MSF), also referred to as "360-degree evaluation," has emerged as an important approach for assessing professional competence, behaviors, and attitudes in the workplace.[4]

Although early attempts at the development of MSF questionnaires in medicine focused on the assessment of residents in the late 1970s, today MSF tools are being used in North America (in Canada and the United States) and Europe (in the Netherlands and the United Kingdom) across a number of physician specialties.[4] As a self-regulating profession, medicine is accountable for ensuring that physicians are competent in the performance of their clinical roles and duties. To aid regulatory bodies in their efforts to monitor physician practice and patient safety, in the late 1990s, Canada was the first country to introduce an MSF process as a viable approach to assessing physician performance. Typically, this feedback is collected using surveys or questionnaires designed to elicit responses from various respondents (e.g., peers, coworkers, patients) and, in some cases, from the physicians themselves through a corresponding self-assessment version of the measurement instrument. MSF has gained widespread acceptance for evaluation of professionals and is seen as a catalyst for the practitioner to reflect on where change may be required.

MSF originated in industry during a time when the search for competent employees and the reliance on a single supervisor's evaluation was recognized as a restrictive approach to the assessment of a worker's specific abilities.[5,6] Similarly, physicians work with a variety of people (e.g., medical colleagues, consultants, therapists, nurses, coworkers) who are able to provide a better assessment and contextually based understanding of physician performance than any single person could. In MSF, physicians may complete a self-assessment instrument and receive feedback from a number of medical colleagues (peers), in-training supervisors or preceptors, nonphysician coworkers (e.g., nurses, psychologists, pharmacists), as well as their own patients.[7] Different respondents focus on characteristics of the physician that they can assess (e.g., patients are not expected to assess a physician's clinical expertise) and together provide a more comprehensive evaluation than what could be derived by any one source alone.[8]

Please see the end of this article for information about the authors.

Correspondence should be addressed to Dr. Donnon, Medical Education and Research Unit, G13 Health Medical Research Building, Faculty of Medicine, University of Calgary, 3330 Hospital Dr., NW, Calgary, AB Canada, T2N 4N1; telephone: (403) 210-9682; fax: (403) 210-7507; e-mail: tldonnon@ucalgary.ca.

Acad Med. 2014;89:00–00.
First published online
doi: 10.1097/ACM.0000000000000147

Supplemental digital content for this article is available at http://links.lww.com/ACADMED/A185.

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

MSF is gaining acceptance and credibility as a means of providing doctors with relevant information about their practice to help them monitor, develop, maintain, and improve their competence. MSF has focused on clinical skills, communication, collaboration with other health care professionals, professionalism, and patient management.[9] Accordingly, the purpose of the present study was to conduct a systematic review of the published, peer-reviewed research on the different types of MSF instruments used to assess physicians' performance on clinical and nonclinical skills and to investigate the evidence for reliability, generalizability, validity, and feasibility of this assessment approach.

Method

Selection of studies

We conducted a systematic review of the research on MSF published from 1975 to January 2013 using the following databases: MEDLINE, PubMed, EMBASE, CINAHL, PsycINFO, and the Cochrane Database of Systematic Reviews. We identified initial search terms to pilot from practical guides and a handbook on MSF.[4,5] The search was limited to English language, peer-reviewed journals, using the terms "multisource-feedback" and "360 degree evaluation" to identify MSF-related studies. We combined these terms with others to capture physician-related assessments: "assessment of physician competencies," and "assessment of physician professionalism," "assessment of physician in practice." We also manually searched from the reference lists of relevant studies.

Eligibility criteria

Studies were included if they (1) used one or more MSF instruments (e.g., feedback from self, colleague, coworker, and/or patient) to assess physician or resident performance in practice; (2) described the MSF instrument or its design; (3) reported psychometric evidence of the instrument(s) in the form of reliability, generalizability, and/or feasibility (administration) of collecting the feedback data; (4) provided evidence of construct and/or criterion-related validity (predictive/concurrent); and (5) were published in an English-language, peer-reviewed journal. We excluded studies if (1) the MSF instrument was used to assess medical students or nonphysician health professionals (i.e., nurses, occupational or respiratory therapists, chiropractors, etc.), and (2) they failed to provide adequate information about the psychometrics of the MSF instrument (reliability and validity). For example, Violato and Lockyer[10] compared mean self and peer MSF ratings between three different specialties, Sinclair et al[11] focused on the issue of patient reliability using the SHEFFPAT questionnaire, and Noonan et al[12] provided information on the test–retest reliability of an MSF instrument, but all three of these studies failed to provide an analysis on the validity of the MSF instruments, so they were excluded. Although the studies included in this systematic review are based on the completion of MSF questionnaires by various assessors, the quality of the studies are considered to be "high" for this type of research, as each study needed to provide evidence of both reliability and construct (or criterion-related) validity to be included.

Data selection and abstraction

To address concerns of bias, we conducted a comprehensive search using strict selection criteria based on rigorous interrater reliability. Each article in the present study was reviewed and coded by two authors (T.D. and A.A.) independently; initially, titles and abstracts were screened before full-text articles were assessed for eligibility (see Figure 1). All four authors independently reviewed all full-text articles until 100% agreement was achieved. Once articles were identified for inclusion, the following information was extracted: the name of the MSF instrument (if a specific name was not provided for the MSF instrument, the generic terms "360-degree evaluation" or "multisource feedback" were used), specialty of physician participants, number of participants, assessor type, construct/factors assessed by the MSF instrument, administration/feasibility issues, mean number of raters per assessor type (response percentage), reliability/generalizability/intraclass correlation coefficients, and analysis of construct and criterion-related validity.

Results

As shown in Figure 1, the review of 96 full-text studies resulted in a total of 43 peer-reviewed articles on physician MSF (see Supplemental Digital Table 1, http://links.lww.com/ACADMED/A185).[7,13–54] Although there were a variety of MSF instruments used in the studies included, the frequency with which they were used was as follows: the Physician Assessment Review (PAR) process (Canada, $n = 13$; Netherlands, $n = 1$), the Sheffield Peer Review Assessment Tool (SPRAT) process (UK, $n = 6$), multiple MSF instruments from the United States ($n = 14$), other UK-related instruments ($n = 4$), and three separate instruments from other countries (China, Denmark, and Taiwan).

Specialty of physicians assessed using MSF

There were a number of MSF studies that assessed physicians across multiple specialties ($n = 10$). In a study of the psychometrics of the PAR MSF instruments, for example, Hall et al[13] evaluated the results from 308 physicians from multiple specialties in Alberta. With respect to specific physician practices, there were MSF studies for each of the following specialties: family medicine ($n = 5$), pediatrics ($n = 5$), internal medicine ($n = 5$), surgery ($n = 4$), obstetrics–gynecology ($n = 3$), psychiatry ($n = 3$), anesthesia ($n = 2$), and one each for emergency medicine, pathology/laboratory medicine, histopathology, radiology, and physical medicine and rehabilitation.

MSF assessors and length of questionnaires

In MSF with physicians, information can come from a variety of sources (i.e., peers or medical colleagues, including supervisors and preceptors; coworkers, such as nurses and other allied health professionals; patients and their families; and self-assessment). In 38 (91%) of the studies, the use of an MSF instrument was completed by the physicians' peers or medical colleagues. In most studies, however, assessments were also obtained from nonphysician coworkers ($n = 32$; 74%), patients and/or their families ($n = 23$; 53%), and self-assessments ($n = 22$; 51%).

The MSF questionnaires varied greatly in the number of items depending on the assessor: 4 to 57 items for self-assessment, 4 to 60 items for peer or medical colleague, 4 to 60 items for coworkers, and 3 to 49 items for patient questionnaires. The PAR studies used

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

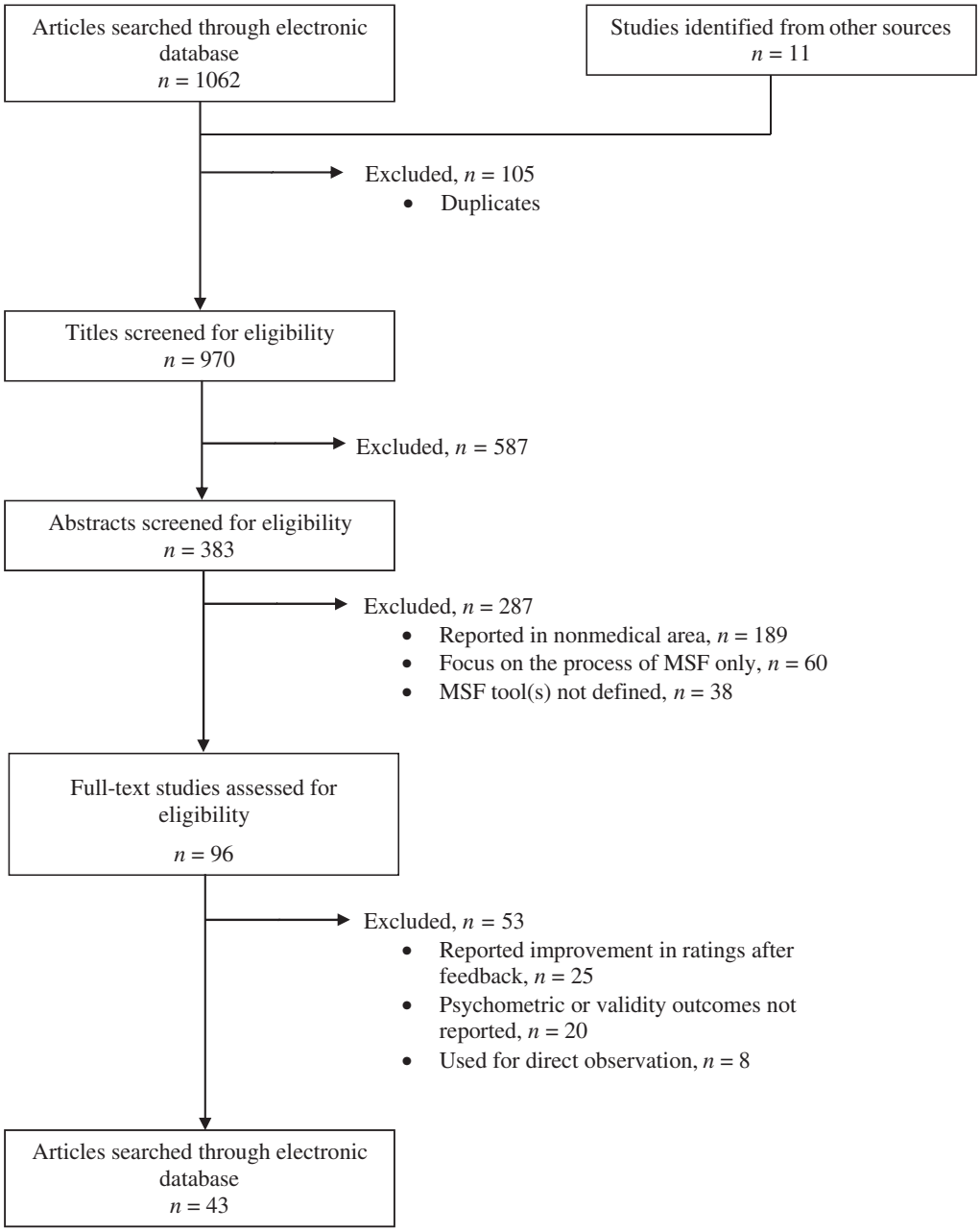

Figure 1 Selection of studies for a systematic review of studies published from 1975 to January 2013 to investigate the reliability, generalizability, validity, and feasibility of multisource feedback (MSF) for the assessment of physicians.

a variety of MSF instruments for each of the assessors, with the number of items (depending on specialty) ranging from 11 to 40 items for the patient, 12 to 22 for the coworker, 22 to 39 for the medical colleague, and 21 to 39 for the self-assessment instrument. The SPRAT uses the same 24-item MSF instrument for medical colleagues and coworkers, although modified versions for histopathology (21-item PATH-SPRAT),[27] junior residents (16-item mini-PAT),[28] and patients (13-item SHEFFPAT)[29] have been introduced. In two studies, medical students were also involved in the MSF process and completed the same 10- or 12-item instrument that medical colleagues, coworkers, and patients used.[39,45]

Constructs/domains assessed

As shown in Supplemental Digital Table 1, http://links.lww.com/ACADMED/A185, a number of constructs were measured using MSF: professionalism, clinical competence, communication, manager, and interpersonal relationship. All of the authors achieved consensus for these five main category domains that, in general, were found to be based on existing constructs or examples of items provided from the included studies. "Professionalism," for example, consisted of a variety of measures of psychosocial skills, professional management/responsibilities, humanistic qualities, compassion, attitude, teaching, and professional development. "Clinical competence" included items that assessed clinical care, good medical practice, patient care, safe practice, clinical performance, clinical knowledge, critical thinking, diagnosis, and management of complex problems. Items connected

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

to the "communication," "interpersonal relationship," and "manager" constructs were grouped and categorized similarly. For example, items that were written "Communicates effectively with patients" or "Communicates effectively with other health care professionals" were clearly associated with the communication category, "Collaborates with medical colleagues" was associated with the interpersonal relationship category, and "Manages health care resources efficiently" was associated with the manager category.[13]

General information on process, administration, and/or feasibility

Each of the 42 studies provided general information about the findings of their study with comments on the process, administration, and/or feasibility (see Supplemental Digital Table 1, http://links.lww.com/ACADMED/A185). For example, general information comments emphasized how studies' psychometric results provided support for the MSF process, how the instrument was able to be administered to the various participants in an efficient manner, and/or how the authors used a feasible method to collect multiple performance measures of physicians in practice. Researchers have acknowledged that the MSF instruments are effective when used in triangulation with patients, coworkers, and medical colleagues in conjunction with the physician's self-assessment.[7] The authors of some studies recognized that the feedback provided to physicians regarding their performance on key competencies has the potential to initiate changes in practice.[14] There was an initial PAR study that considered MSF to be feasible as a function of the estimated cost per physician, but it was suggested that the MSF on the physician be readministered every five years.[13] In a subsequent PAR study, family medicine physicians were assessed and then reassessed after five years (i.e., Time 1 and Time 2), providing evidence of measurement stability; however, the incorporation of feedback by the physicians was limited.[20,21] In PAR-related studies, the administration of the MSF process was found to be feasible and adaptable for a variety of specialties (e.g., pediatrics,[19] surgery,[14] emergency medicine,[17] family medicine,[20] psychiatry[22]) and potentially for use in other countries.[24] Although the SPRAT originated with the use of a common 24-item MSF instrument for medical colleagues and coworkers in pediatrics, modified versions of the peer-review assessment instruments has also been used with multiple specialities.[26–31] In 2008, the study by Crossley et al[29] introduced a 13-item patient MSF instrument (the Sheffield Patient Assessment Tool) that, in a subsequent study by Archer and McAvoy,[31] failed to show that patients were able to identify doctors in potential difficulty.

Reliability and generalizability of MSF instruments

The reliability of the various MSF instruments was reported in 26 (62%) of the studies included in this systematic review. Reliability coefficients are reported typically as Cronbach alpha (α) and reflect the internal consistency of the items. MSF instruments should have an $\alpha \geq 0.90$, which was typically achieved in PAR-related studies for the medical colleague (0.89–0.99), coworker (0.91–0.96), and patient (0.93–0.99) instruments. Although only one of the SPRAT studies included a combined medical colleague and coworker reliability coefficient ($\alpha = 0.98$),[28] the standard error of measurement (SEM) was calculated for 5 of the 6 included studies. In general, to achieve an SEM of $\pm$ 0.40 with the combined SPRAT, a minimum of eight raters is required.

Using generalizability analyses, generalizability coefficients (Ep^2) were derived in 17 studies (40%). Ep^2 provides a measure of the dependability of the MSF instruments as a function of the various factors that can influence the physicians' ratings. The coefficients for the medical colleague instrument ranged from $Ep^2 = 0.61$ to 0.88, for the coworker instrument ranged from 0.56 to 0.87, and for the patient instrument ranged from 0.65 to 0.85. In four studies, the intraclass correlation coefficient (ICC) was calculated as a way to determine the consistency in ratings across the evaluators and ranged from 0.45 to 0.90 (suggesting that the ratings obtained from the various evaluators were moderate to highly consistent).

As shown in Supplemental Digital Table 2, http://links.lww.com/ACADMED/A185, assessment of physician performance was based on the completion of the MSF instruments by various numbers of multiple stakeholders. In summary, most of the instruments required a minimum of 8 medical colleagues, 8 coworkers, and 25 patients to achieve adequate reliability and generalizability coefficients of $\alpha \geq 0.90$ and $Ep^2 \geq 0.80$, respectively.

Construct and criterion-related validity

To be included in this systematic review, a study had to provide evidence of either construct and/or criterion-related validity (predictive/concurrent). In 28 (67%) of the studies, evidence for the construct validity of the MSF instrument used was provided through exploratory factor analyses (principal component). As we have seen, each of the MSF instruments was found to assess a variety of constructs based on the particular instrument used (i.e., PAR, SPRAT, other) or the respondent (i.e., medical colleague, coworker, patient).

Further evidence of construct validity was provided through analyses that showed (1) measures of mean difference ratings between respondent groups (i.e., mean ratings from patients and coworkers are consistently higher than medical colleagues' and are lowest on self-assessments), (2) improvement in performance ratings from Time 1 to Time 2 (i.e., mean ratings are consistently higher compared with an earlier assessment period, indicating an expected improvement in practice over time), (3) consistently higher ratings given to advanced trainees by year of program (i.e., increase in mean ratings as residents gain clinical experience from year to year of an in-training program), and (4) higher ratings for younger practitioners than older ones (i.e., higher mean ratings are generally given to young practitioners who have been educated to be more conscious of MSF domain measures than practitioners that have been in practice for a greater number of years). In 30 (71%) of the studies, evidence of construct validity was supported with findings that patients, followed by coworkers, tended to rate physicians more positively than did residents, who were more positive still than faculty and consultant raters.

Criterion-related validity was indicated in some studies where positive correlations were found between the

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

MSF instruments/measures (concurrent validity), and between MSF ratings and other assessment instruments/measures (predictive or concurrent validity). As reported in Risucci et al,[33] there was strong concurrent validity for the medical colleague MSF questionnaire where supervisor and peer mean ratings on the same measures of physician performance correlated at $r = 0.92$ ($P < .001$). The PATH-SPRAT total aggregated score, for example, was found to correlate at $r = 0.48$ ($P < .001$) with histopathology residents' performance on an objective structured practice examination.[27]

Discussion

In a review of the MSF instruments included in this systematic review, there appears to be agreement that the administration of a 360-degree evaluation of physicians in practice from a variety of specialties is feasible from self-assessment, medical colleague, coworker, and patient perspectives. Most studies that provide evidence of reliability, generalizability, and validity (construct and criterion-related) are from the PAR process in Canada and the SPRAT instruments used in the United Kingdom, where the longitudinal and multistudy nature of the MSF research on physician performance has been in progress for 16 and 8 years, respectively. Although there are a number of U.S. MSF studies (14), each of these articles focused on the use of a new MSF instrument or a modified version of an existing instrument/evaluation guideline (see Supplemental Digital Table 1, http://links.lww.com/ACADMED/A185).

In general, physician performance assessment with MSF instruments employed a minimum of 8 medical colleagues, 8 coworkers, and 25 patients to achieve reliability and generalizability coefficients of $\alpha \geq 0.90$ and $Ep^2 \geq 0.80$, respectively. Although a variety of constructs were assessed, there were five key domains identified across the MSF instruments: (1) professionalism, (2), clinical competence, (3) communication, (4) manager, and (5) interpersonal relationships. The majority of the studies provided evidence of the construct validity of the MSF instruments used by conducting a principal component factor analysis or comparing mean rating scores between rater groups. Although typically patients tended to rate physicians most positively, followed by coworkers, resident peers, faculty, and consultant evaluators, we were interested to see that Lockyer et al[16] found that self-assessments were higher than peers' assessments in a general practice sample of international medical graduates. While the construct validity of MSF questionnaires may be found within a particular discipline (e.g., family medicine, internal medicine, surgery), many authors acknowledged that measures of various competencies or constructs are a function of the specialization assessed (i.e., the percentage of variance associated with measures of patient management, clinical assessment, communication, and/or professional development was found to vary across specialties).[10,15,30,34] For example, Lockyer and Violato[15] found in a principal component factor analysis of a medical colleague MSF questionnaire that the resulting four-factor solution accounted for 73.4% of the variance for internal medicine physicians, 70% for psychiatrists, and only 67.6% for pediatricians.

Although our systematic review was rigorous, there are limitations to the present study. First, there is heterogeneity in the MSF instruments used and the number of items employed to measure the various constructs identified. Accordingly, the identification of a single best MSF instrument is difficult and context/specialty-specific. Second, the feasibility of using MSF is based primarily on the reported response rate percentages but does not typically include costs and administration concerns in the assessment of physician performance. Third, variability in the reporting of reliability (i.e., generalizability, intraclass correlation) and validity (i.e., construct- and criterion-related) measures, while supportive of the MSF process, were difficult to combine consistently between studies. Finally, our search was limited to English-language peer-reviewed journal articles and may not reflect MSF processes in other countries or those currently in use but not published.

In summary, MSF where various assessors (self, peers, coworkers, and patients) provide assessment of physicians' performance on various domains (clinical and nonclinical) is reliable, valid, and feasible. As indicated above, there exists a substantial body of rigorous and consistent research on the PAR and SPRAT programs demonstrating that the use of MSF will continue to play an important role in the formative and potentially summative assessment of physician performance in practice. Future research should focus on consolidating measures of competence domains between and within physician specialties, while taking into consideration issues related to the establishment of an MSF process at local and national levels.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

Dr. Donnon is associate professor, Medical Education and Research Unit, Department of Community Health Sciences, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada.

Dr. Al Ansari is director of training and development, Department of Medical Education, Faculty of Medicine, Bahrain Defense Force Hospital, Riffa, Bahrain.

Dr. Al Alawi is a faculty member, Department of Family Medicine, Faculty of Medicine, Bahrain Defense Force Hospital, Riffa, Bahrain.

Dr. Violato is professor, Medical Education and Research Unit, Department of Community Health Sciences, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada.

References

1 Kohn LT, Corrigan JM, Donaldson MS, eds. Too Err Is Human: Building a Safer Health System. Washington, DC: National Academy Press; 1999.
2 Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA. 2002;287:226–235.
3 Bandiera G, Sherbino J, Frank JR. The CanMEDS Assessment Tool Handbook. An Introductory Guide to Assessment of the CanMEDS Competencies. Ottawa, Ontario, Canada: Royal College of Physicians and Surgeons of Canada; 2006.
4 Lockyer J, Clyman S. Multisource feedback (360-degree evaluation). In: Holmboe ES, Hawkins RE, eds. Practical Guide to the Evaluation of Clinical Competence. Philadelphia, Pa: Mosby; 2008.
5 Bracken DW, Timmreck CW, Church AH. Introduction: A multisource feedback process model. In: Bracken DW, Timmreck CW, Church AH, eds. The Handbook of Multisource Feedback: The Comprehensive Resource for Designing and Implementing MSF Processes. San Francisco, Calif: Jossey-Bass; 2001:3–14.
6 Bracken DW, Church AH. Advancing the state of the art of 360-degree feedback: Guest editors' comments on the research and practice of multi rater assessment methods. Group Org Manag. 1997;22:149–161.

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

7 Violato C, Marini A, Towes J, et al. Feasibility and psychometric properties of using peers, consulting physicians, co-workers, and patients to assess physicians. Acad Med. 1997;72:82–84.
8 Sala F, Dwight S. Predicting executive performance with multi-rater surveys: Whom you ask makes a difference. J Consult Psych Res Pract. 2002;54:166–172.
9 Fidler H, Lockyer J, Violato C. Changing physicians' practices: The effect of individual feedback. Acad Med. 1999;74:702–714.
10 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.
11 Sinclair AM, Gunendran T, Archer J, et al. Re-certification for urologists: Is the SHEFFPAT questionnaire valid for assessing clinicians' "relationships with patients"? Br J Med Surg Urol. 2009;2:100–104.
12 Noonan CL, Monagle J, Castanelli D. Development of a multi-source feedback tool for consultant anaesthetist performance. Aust Health Rev. 2011;35:141–145.
13 Hall W, Violato C, Lewkonia R, et al. Assessment of physician performance in Alberta: The Physician Achievement Review. CMAJ. 1999;161:52–57.
14 Violato C, Lockyer J, Fidler H. Multisource feedback: A method of assessing surgical practice. BMJ. 2003;326:546–548.
15 Lockyer JM, 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–S8.
16 Lockyer J, Blackmore D, Fidler H, et al. A study of multi-source feedback system for international medical graduates holding defined licences. Med Educ. 2006;40:340–347.
17 Lockyer JM, Violato C, Fidler H. The assessment of emergency physicians by a regulatory authority. Acad Emerg Med. 2006;13:1296–1303.
18 Lockyer J, Violato C, Fidler H. A multi source feedback program for anesthesiology. Can J Anesth. 2006;53:33–39.
19 Violato C, Lockyer JM, Fidler H. Assessment of pediatricians by a regulatory authority. Pediatrics. 2006;117:796–802.
20 Lockyer JM, Violato C, Fidler HM. What multisource feedback factors influence physician self-assessments? A five-year longitudinal study. Acad Med. 2007;82(10 suppl):S77–S80.
21 Violato C, Lockyer JM, Fidler H. Changes in performance: A 5-year longitudinal study of participants in a multi-source feedback programme. Med Educ. 2008;42:1007–1013.
22 Violato C, Lockyer JM, Fidler H. Assessment of psychiatrists in practice through multisource feedback. Can J Psychiatry. 2008;53:525–533.
23 Lockyer J, Violato C, Fidler H, et al. The assessment of pathologists/laboratory medicine physicians through a multisource feedback tool. Arch Pathol Lab Med. 2009;133:1301–1308.
24 Overeem K, Wollersheim H, Arah OA, et al. Evaluation of physicians' professional performance: An iterative development and validation study of multisource feedback instruments. BMC Health Serv Res. March 26, 2012;12:80.
25 Lockyer J, Violato C, Wright B, et al. Long-term outcomes for surgeons from 3- and 4-year medical school curricula. Can J Surg. 2012;55:S1–S5.
26 Archer JC, Norcini J, Davies HA. Use of SPRAT for peer review of paediatricians in training. BMJ. 2005;330:1251–1253.
27 Davies H, Archer J, Bateman A, et al. Specialty-specific multi-source feedback: Assuring validity, information training. Med Educ. 2008;42:1014–1020.
28 Archer J, Norcini J, Southgate L, et al. Mini-PAT (Peer Assessment Tool): A valid component of a national assessment programme in the UK? Adv Health Sc Educ. 2008;13:181–192.
29 Crossley J, McDonnell J, Cooper C, et al. Can a district hospital assess its doctors for re-licensure? Med Educ. 2008;42:359–363.
30 Archer J, McGraw M, Davies H. Assuring validity of multisource feedback in a national programme. Postgrad Med J. 2010;86:526–531.
31 Archer JC, McAvoy P. Factors that might undermine the validity of patient and multi-source feedback. Med Educ. 2011;45:886–893.
32 DiMatteo MR, DiNicola DD. Sources of assessment of physician performance: A study of comparative reliability and patterns of intercorrelation. Med Care. 1981;19:829–842.
33 Risucci DA, Tortolani AJ, Ward RJ. Ratings of surgical residents by self, supervisors and peers. Surg Gynecol Obstet. 1989;169:519–526.
34 Ramsey PG, Wenrich MD, Carline JD, et al. Use of peer ratings to evaluate physician performance. JAMA. 1993;269:1655–1660.
35 Wenrich MD, Carline JD, Giles LM, et al. Ratings of the performances of practicing internists by hospital-based registered nurses. Acad Med. 1993;68:680–687.
36 Thomas PA, Gebo KA, Hellmann DB. A pilot study of peer review in residency training. J Gen Intern Med. 1999;14:551–554.
37 Lipner RS, Blank LL, Leas BF, et al. The value of patient and peer ratings in recertification. Acad Med. 2002;77(10 suppl):S64–S66.
38 Davis JD. Comparison of faculty, peer, self, and nurse assessment of obstetrics and gynecology residents. Obstet Gynecol. 2002;99:647–651.
39 Joshi R, Ling FW, Jaeger J. Assessment of a 360-degree instrument to evaluate residents' competency in interpersonal and communication skills. Acad Med. 2004;79:458–463.
40 Wood J, Collins J, Burnside ES, et al. Patient, faculty, and self- assessment of radiology resident performance: A 360-degree method of measuring professionalism and interpersonal/communication skills. Acad Radiol. 2004;11:931–939.
41 Wood L, Wall D, Bullock A, et al. "Team observation": A six-year study of the development and use of multi-source feedback (360-degree assessment) in obstetrics and gynecology training in the UK. Med Teach. 2006;28:e177–e184.
42 Brinkman WB, Geraghty SR, Lanpher BP, et al. Effect of multisource feedback on resident communication skills and professionalism. Arch Pediatr Adolesc Med. 2007;161:44–49.
43 Allerup P, Aspegren K, Ejlersen E, et al. Use of 360-degree assessment of residents in internal medicine in a Danish setting: A feasibility study. Med Teach. 2007;29:166–170.
44 Pollock RA, Donnelly MB, Plymale MA, et al. 360-degree evaluations of plastic surgery resident Accreditation Council for Graduate Medical Education competencies: Experience using a short form. Plast Reconstr Surg. 2008;122:639–649.
45 Massagli TL, Carline JD. Reliability of a 360-degree evaluation to assess resident competence. Am J Phys Med Rehabil. 2007;86:845–852.
46 Lelliott P, Williams R, Mears A, et al. Questionnaires for 360-degree assessment of consultant psychiatrists: Development and psychometric properties. Br J Psychol. 2008;193:156–160.
47 Campbell JL, Richards SH, Dickens A, et al. Assessing the professional performance of UK doctors: An evaluation of the utility of the General Medical Council patient and colleague questionnaires. Qual Saf Health Care. 2008;17:187–193.
48 Meng L, Metro DG, Patel RM. Evaluating professionalism and interpersonal and communication skills: Implementing a 360-degree evaluation instrument in an anesthesiology residency program. J Grad Med Educ. 2009;1:216–220.
49 Campbell J, Narayanan A, Burford B, et al. Validation of a multi-source feedback tool for use in general practice. Educ Prim Care. 2010;21:165–179.
50 Chandler N, Henderson G, Park B, et al. Use of a 360-degree evaluation in the outpatient settings: The usefulness of nurse, faculty, patient/family, and resident self-evaluation. J Grad Med Educ. 2010;10:430–434.
51 Yang YY, Lee FY, Hsu HC, et al. Assessment of first-year post-graduate residents: Usefulness of multiple tools. J Chin Med Assoc. 2011;74:531–538.
52 Wall D, Singh D, Whitehouse A, et al. Self-assessment by trainees using self-TAB as part of the team assessment of behavior multisource feedback tool. Med Teach. 2012;34:165–167.
53 Qu B, Zhao YH, Sun BZ. Assessment of resident physicians in professionalism, interpersonal and communication skills: A multisource feedback. Int J Med Sci. 2012;9:228–236.
54 Wright C, Richards SH, Hill JJ, et al. Multisource feedback in evaluating the performance of doctors: The example of the UK General Medical Council patient and colleague questionnaires. Acad Med. 2012;87:1668–1678.

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.