

Feasibility of Implementing a Standardized Multisource Feedback Program in the Graduate Medical Education Environment

MARGARET RICHMOND, MS
COLLEEN CANAVAN, MS
MATTHEW C. HOLTMAN, PhD
PETER J. KATSUFRAKIS, MD, MBA

Abstract

Background Multisource feedback (MSF) is emerging as a central assessment method for several medical education competencies. Planning and resource requirements for a successful implementation can be significant. Our goal is to examine barriers and challenges to a successful multisite MSF implementation, and identify the benefits of MSF as perceived by participants.

Methods We analyzed the 2007–2008 field trial implementation of the Assessment of Professional Behaviors, an MSF program of the National Board of Medical Examiners, conducted with 8 residency and fellowship programs at 4 institutions. We use a multimethod analysis that draws on quantitative process indicators and qualitative participant experience data. Process indicators include program attrition, completion of implementation milestones, number of participants at each site, number of MSF surveys assigned and completed, and adherence to an experimental rater training protocol. Qualitative data

include communications with each program and semistructured interviews conducted with key field trial staff to elicit their experiences with implementation.

Results Several implementation challenges are identified, including communication gaps and difficulty scheduling implementation and training workshops. Participant interviews indicate several program changes that should enhance feasibility, including increasing communication and streamlining the training process.

Conclusions Multisource feedback is a complex educational intervention that has the potential to provide users with a better understanding of performance expectations in the graduate medical education environment. Standardization of the implementation processes and tools should reduce the burden on program administrators and participants. Further study is warranted to broaden our understanding of the resource requirements for a successful MSF implementation and to show how outcomes change as MSF gains broader acceptance.

Editor's Note: The online version of this article contains the interview guide used in this study.

Background

Multisource feedback (MSF) is an assessment approach that uses input from peers and colleagues to gather information about an individual's behavior in the workplace.¹ This information is then aggregated and provided to

the individual as feedback. Multisource feedback has been recommended by the Accreditation Council for Graduate Medical Education (ACGME) as a key method for assessing several of the competencies, including professionalism, and interpersonal and communication skills.² Reliability, validity, and feasibility are central considerations in selecting assessment tools.³

Feasibility in assessment considers costs, logistical considerations, and user acceptance.^{4,5} Despite its importance, feasibility is an underexamined construct that is given only superficial attention in many validity studies. There is a small base of literature supporting feasibility of MSF in the medical training environment.^{6,7} However, only a few studies^{8–14} address use of MSF in graduate medical education, including one of the early implementations of the MSF program described in this study.¹⁵ Authors note the time burden associated with survey completion^{8,11–14} and often low survey completion rates,¹⁵ participants' perceptions of program feasibility,^{14,15} or report that a given site or rater group was excluded on feasibility grounds without providing specific detail.^{9,12}

Margaret Richmond, MS, is Project Manager at the National Board of Medical Examiners; **Colleen Canavan, MS**, is a Program Associate at the National Board of Medical Examiners; **Matthew C. Holtman, PhD**, is a Senior Manager at IFC International; and **Peter J. Katsufraakis, MD, MBA**, is a Vice President at the National Board of Medical Examiners.

Funding: The authors report no external funding source.

Corresponding author: Margaret Richmond, MS, National Board of Medical Examiners, 3750 Market Street, Philadelphia, PA 19104, 215.590.9504, mrichmond@nbme.org

Received May 24, 2010; revisions received February 28, 2011 and July 28, 2011; accepted July 22, 2011.

DOI: <http://dx.doi.org/10.4300/JGME-D-10-00088.1>

The organizational problems inherent in planning and carrying out an MSF program are only rarely addressed explicitly.^{1,15} Furthermore, few studies report on the feasibility issues of MSF implementations that span more than 1 site.¹⁶ If MSF is to achieve its potential in graduate medical education, it will be necessary to understand and find solutions to the implementation barriers.

Our study begins to address this gap in knowledge through an analysis of outcomes associated with a 2007–2008 multisite field trial of the Assessment of Professional Behaviors (APB). The APB is an MSF program developed by the National Board of Medical Examiners (NBME) to provide physicians, medical students, residents, and fellows with feedback on the professional behaviors that are essential to the safe, effective, and ethical practice of medicine. The APB program is built around a standardized MSF survey instrument developed with input from experts in medical education, medical professionalism, psychometrics, survey design, and industrial organizational psychology. The purpose of this study was to examine (1) program implementation, completion, and attrition; (2) the communication challenges associated with implementation; (3) reasons for implementation delays; and (4) benefits of the MSF program as perceived by participants.

Methods

A description of the MSF instrument used by all participating sites is provided in TABLE 1. We chose behaviors for inclusion in the instrument based on surveys of potential respondents at the participating sites before program implementation; these surveys asked about the behavioral items’ observability, clarity, and importance in the graduate medical education environment. Participating sites obtained local Institutional Review Board (IRB) approval for this study, and each site was responsible for obtaining consent from participants as required by their IRB.

What was known
Multisource feedback has shown promise for competencies not well assessed using traditional methods.

What is new
A multisite study found implementation challenges due to communication and scheduling problems and need for faculty development.

Limitations
Small, potentially nonrepresentative sample; significant attrition.

Bottom line
Standardizing implementation and adapting faculty development to time constraints may increase/preserve participation.

The APB program includes a leadership orientation, participant orientation, feedback provider training, delivery of feedback reports, and feedback sessions. As part of the 2007–2008 field trial, a rater training experiment was also performed during the course of implementation. Each ABP milestone, along with the initial timeline of the APB field trial implementation, is described in TABLE 2.

Residency and fellowship programs were eligible to participate in the field trial if they were located in the northeastern or midwestern United States and based at an institution where multiple programs were using 1 major residency management and evaluation system. The NBME contacted the designated institutional officials (DIOs) of institutions meeting these criteria to invite their participation in the field trial. Interested DIOs were asked to distribute field trial information to their program directors, including a request for statements of interest.

Programs were responsible for orienting all participants to the APB program. Potential raters included attending physicians, faculty, residents, fellows, nurses, and other staff. Subjects of the evaluation were primarily residents

TABLE 1 ASSESSMENT OF PROFESSIONAL BEHAVIORS MULTISOURCE FEEDBACK INSTRUMENT: DESCRIPTION/SAMPLE ITEMS		
Section	Description	Sample Item(s)
Relational	4 items regarding the amount and types of interaction the rater had with the observee during the evaluation period	I had direct contact with or observed this person...(less than once a week, 1–3 times a week, 4–7 times a week, more than 7 times a week).
Behavioral	25 specific, observable behaviors rated on a 5-point frequency scale	▪ Maintains composure during difficult interactions ▪ Responds promptly when paged or called ▪ Acknowledges limits of own knowledge or ability
Global	1 global evaluation item rated on a 5-point agreement scale	I would like to have this person on my team.
Narrative	2 free-text comment spaces to identify behaviors exhibited by the observee that are praiseworthy and those that need improvement	Please use the comment box below to record specific examples of professional behavior exhibited by this person that merit praise or commendation.

TABLE 2 DESCRIPTION OF ASSESSMENT OF PROFESSIONAL BEHAVIORS (APB) PROGRAM FIELD TRIAL MILESTONES

Milestone	Audience	Description
Leadership orientation (Sept. 2007–Oct. 2007)	DIOs, program directors, directors of nursing, other institutional leaders	To provide an overview of the APB program for institutional and program leaders
		Addressed APB purpose, implementation stages and recommended resources, strategies for generating commitment and buy-in
		Presented by NBME staff onsite at participating institutions
		Time allotted: 1 to 1½ hours
Participant orientation (Oct. 2007–Dec. 2007)	All individuals participating in the APB (raters, observees, feedback providers)	To familiarize participants with the APB implementation and to generate buy-in to the process; provided overview of APB, expectations for participation, desired outcomes
		Presented by program leadership with delivery method of their choice
		Time allotted: one-half to 1 hour
Rater training (Oct. 2007–Dec. 2007)	All raters randomly assigned to training interventions	To educate raters on effective use of the APB instrument
		Addressed APB behaviors and rating scale, common rater errors and challenges, writing effective feedback comments
		Online asynchronous OR onsite workshop presented by NBME-hired consultant and NBME staff
		Time allotted: 2 hours
Feedback provider Training (Feb. 2008–May 2008)	Faculty/mentors conducting feedback sessions with recipients	To provide guidance to faculty on conducting feedback sessions
		Addressed APB background, characteristics of effective feedback, suggested sequence of events for a feedback session
		Onsite workshops presented by NBME staff and NBME-hired consultant. Time allotted: 2 hours
Feedback reports (Feb. 2008–June 2008)	Observees, feedback providers, program leadership	To display aggregate feedback for individual observees evaluated in the APB program; generated by NBME and delivered to participants via evaluation software
Feedback sessions (Feb. 2008–June 2008)	Observees, feedback providers	To individually review feedback reports with observees and generate plans for behavioral improvement
		Delivery determined by program leadership
		Time allotted: suggested 15 minutes per session

Abbreviations: DIOs, designated institutional officials; NBME, National Board of Medical Examiners.

and fellows. As part of a randomized, controlled rater training experiment, all programs were asked to submit a list of individuals participating as raters. The NBME randomly assigned the raters to web-based training, workshop training, or no training. The NBME suggested that training be delivered approximately 2 weeks before participants were to be assigned the MSF instrument. The NBME worked with the programs and DIOs to schedule the workshop training sessions delivered by NBME staff and contracted training consultants. Instructions for accessing the web-based training were delivered locally through the program's evaluation system at about the same time as the live workshops.

Upon completion of the rater training, sites were responsible for assigning the MSF instrument. The instrument was loaded by the NBME into the evaluation software (E*Value, Advanced Informatics, Minneapolis,

MN) and made available through the participating programs' accounts so that local staff could administer the MSF instrument on a locally defined schedule. Participating programs determined who would observe and rate whom and how often. After collecting MSF data for a period of time, the programs were to notify the NBME when they wanted to receive their feedback reports. Through a data exchange system between NBME and the evaluation software vendor, NBME generated and delivered reports to the individual observees (ie, those being observed and evaluated) through the evaluation system.

To convey the breadth of the field trial experience and the extent to which the reality of implementation departed from our expectations, we use mixed quantitative and qualitative methods¹⁷ to analyze both process and outcome data. Process indicator data included program attrition, timing of key implementation milestone completion,

TABLE 3 PARTICIPATION AT 8 FIELD TRIAL SITES

Program	Observees, No.	Surveys Sent, No.	Surveys Completed, No.	Completion Rate, %
General internal medicine residency	129	1335	1126	84.3
Pediatrics residency	88	2190	683	31.2
Internal medicine pulmonary diseases/critical care residency	37	162	140	86.4
General psychiatry residency	21	441	58	13.2
Child and adolescent psychiatry residency	11	44	32	72.7
Gastroenterology fellowship	9	36	33	91.7
Rheumatology fellowship	5	88	68	77.3
Anesthesiology residency	3	12	8	66.7
Total	303	4308	2148	65.4

number of participants and number of MSF surveys assigned and completed, and adherence to the rater training experiment protocol.

To further understand the local implementation issues, interviews were conducted with stakeholders from the various sites, including those who completed an implementation cycle (defined as completion of at least 1 round of MSF data collection) and those who dropped out of the field trial before completing a full cycle. A small monetary incentive was offered to interviewees. A total of 28 stakeholders were interviewed, including 6 from sites that did not complete a full cycle. Stakeholders included program champions (those responsible for implementing APB at their site), program coordinators, faculty, residents and fellows, and designated institutional officials. A sample of the guiding questions asked of participants is provided as supplemental material.

Results

Program Implementation, Completion, and Attrition

Among the 27 programs at 5 institutions that submitted letters of interest, significant attrition from the field trial occurred almost immediately. One institution with 5 interested programs withdrew early in the process because the institution's implementation leaders had fundamentally misunderstood the purpose of the field trial, highlighting communication failure as one early barrier to implementation. An additional 10 programs withdrew before administering the MSF instrument, with the most common stated reasons for withdrawal being lack of faculty commitment and program resources for implementation.

During interviews, participants indicated that the time commitment required for rater training was a major obstacle to their participation. This concern was expressed among both programs that continued in the field trial and programs that dropped out. Program leaders also suggested that these delays reduced morale and diminished their motivation to participate in the field trial.

Of the 12 remaining training programs that started administering the instrument, 8 completed a full implementation cycle. TABLE 3 shows that 303 unique observees participated and were rated with 2148 surveys. The average survey completion rate was 65.4%.

Communication Challenges

Postimplementation interviews with 22 key staff members at sites that completed an implementation cycle indicated that uncertainty regarding responsibilities remained at several programs. Communication challenges began during recruitment when the DIOs were identified as the initial point of contact regarding the APB field trial. As a result, many program champions did not immediately appreciate the resource commitment that would be needed. Additionally, while role expectations were communicated by NBME staff during the leadership orientation, staff members who were key to the implementation still were unclear about their responsibilities, in some cases owing to inability to attend the orientation. For example, some program coordinators indicated that they did not understand why they were involved, or did not anticipate the amount of work that would be required. Interview respondents reported that despite their sites' participation in these orientations, little information about the program made its way to local stakeholders before implementation.

TABLE 4 RATER TRAINING EXPERIMENT PARTICIPATION

Experimental Assignment	No Training	Training Completed			
		Web Training	Live Workshop Training	Row Total	Adherence Rate, %
No training	228	0	1	229	99.6
Web training	201	33	1	235	14.0
Live workshop training	136	0	90	226	39.8
No assignment made	6	0	0	6	Not applicable
Total	571	33	92	696	

This is paralleled by the trainers' experience in conducting the workshop rater trainings. Many of the participants did not understand why they were at the workshops, or the purpose of the APB program, suggesting that not all participants had been oriented to the APB program.

Reasons for Implementation Delays

Completion of every significant project milestone was delayed at some sites. Survey administration and feedback provider training were delayed by about 4 months at all sites. Structured feedback sessions were delayed by up to 8 months relative to the intended schedule.

Initial delays resulted from poor coordination of communication between the DIOs, the residency programs, and the NBME at several institutions. For example, 1 institution did not supply a list of participating residency programs until 4 months after leadership training was completed. This made it difficult to determine which programs would be participating in the field trial and thus delayed the scheduling of additional training workshops and the completion of subsequent implementation milestones.

A second reason for implementation delays was the rater training experiment. Sites took between 1 to 4 weeks to compile lists of participants for experimental assignment. Delays also occurred owing to the following:

- Negotiation with the sites over the scope of the experiment;
- Additional work required of the site administrators to schedule the participants into workshop sessions and assign web-based training in the evaluation software;
- An assignment process that was complicated by coordinators' varying familiarity with the software;
- Longer time required of the NBME to prepare the training modules than had been anticipated because of the need to train external consultants to deliver workshops, and because of information technology problems associated with the hosting of video content for the web-based rater training module;

- Workshop scheduling constraints due to the need to coordinate participant and trainer availability.

Despite the assignment of two-thirds of raters to a rater training intervention, uptake of training was minimal (TABLE 4).

Furthermore, the assignment of MSF surveys was sufficiently delayed such that any impact of training on rater behavior would likely have been attenuated.

Benefits of MSF Program as Perceived by Participants

Despite the problems, field trial participants reported a number of benefits of participation. Generally, the problems with rater-based assessments were well known among the participants. Interview respondents indicated that they believed the APB program addressed some of those problems by providing appropriate structure and a clear focus on "professionalism" rather than being an add-on to existing assessments. The behavioral focus of the form was perceived as an improvement over current practice. Respondents also reported that the form was well structured and easy to use. Although participants expressed concern over rater fatigue, with some arguing that the 25-item form was too long, participants also recognized that a significantly shorter form would be too "vague" to be useful for giving feedback to learners.

Respondents recognized the difficulty of delivering feedback effectively and considered the APB to be helpful, and the feedback provider training program was also valued. However, dissemination of feedback provider training was subject to many of the same difficulties affecting dissemination of the other training modules.

Discussion

Our research identified several implementation challenges of the APB field trial, including communication gaps and difficulty scheduling onsite rater training. These challenges contributed to implementation delays, program attrition, and low training uptake. Participants identified several

program changes that should enhance feasibility, including increasing communication and streamlining the training process. Despite the challenges, many participants still found value in implementing the APB program, as it helped to clarify performance expectations around professionalism and enhance the feedback process.

Our findings are consistent with those of previous research efforts that have identified the complexity and importance of communication in implementing an MSF program,¹ and that have found MSF instruments to be of value in defining performance criteria about topics that may not otherwise be explicitly addressed.^{1,15} Our study, which includes analysis across several voluntary implementation sites in the United States, contributes to the limited literature on feasibility of multisite MSF implementation, which previously has been focused in the United Kingdom where MSF assessment is mandated.¹⁶

An important limitation of this study is that it is based on a relatively small and likely nonrepresentative sample of training programs. As described in the analysis, program attrition from the field trial was significant and attributed often to lack of resources for implementation. Sample selection bias is therefore a potential problem and may have produced better outcomes in this study than should be expected in the general population. Additionally, dependence on 1 residency management software system may have yielded results that do not generalize to other systems. The impact of rater training, while potentially important, could not be measured in the current study.

Conclusions

Acceptance by participants is an essential element of MSF feasibility, and true acceptance cannot be established until participants are familiar with the components and requirements. Standardization and clear communication are therefore critical. Our experience underscores the importance of establishing the feasibility and acceptability of a well-defined intervention before attempting experimental research to demonstrate effectiveness.¹⁸ Our findings provide further evidence of the complexities involved in implementing an MSF program. Despite the problems encountered, participants in the APB field trial identified several benefits of the program. Streamlining the implementation process may increase the feasibility and standardize the use of the APB and other MSF interventions among a broad population of training programs. Our research points to some important ACGME program requirements, particularly frequent communication and reduction of barriers (eg, scheduling, technology) to fulfilling training requirements that should lead to better

outcomes for any organization undertaking a new MSF program. Organizational learning seems to be a critical ingredient in the success of MSF.¹⁹ We anticipate that MSF implementation processes will improve over time as programs gain more familiarity with this assessment modality. Further research is warranted to better understand the resource requirements for a successful MSF implementation and to show how outcomes change as MSF gains acceptance by the broader graduate medical education community.

References

- 1 Lockyer JM, Clyman SG. Multisource feedback (360-degree evaluation). In: Holmboe ES, Hawkins RE, eds. *Practical Guide to Evaluation of Clinical Competence*. Philadelphia, PA: Mosby Elsevier; 2008:75–85.
- 2 Accreditation Council for Graduate Medical Education. Common Program Requirements: general competencies. <http://www.acgme.org/outcome/comp/GeneralCompetenciesStandards21307.pdf>. Accessed February 17, 2010.
- 3 Accreditation Council for Graduate Medical Education. *Key Considerations for Selecting Assessment Instruments and Implementing Assessment Systems*. 2010.
- 4 Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ*. 1996;1(1):41–67.
- 5 Wagner D, Lypson ML. Centralized assessment in graduate medical education: cents and sensibilities. *J Grad Med Educ*. 2009;1(1):21–27.
- 6 Ramsey P, Wenrich M, Carline J, Inui T, Larson E, LoGerfo J. Use of peer ratings to evaluate physician performance. *JAMA*. 1993;269(13):1655–1660.
- 7 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(suppl 10):S82–S84.
- 8 Archer JC, Norcini J, Davies HA. Use of SPRAT for peer review of paediatricians in training. *BMJ*. 2005;330:1251–1253.
- 9 Brinkman WB, Geraghty SR, Lanphear BP, Khoury JC, Gonzalez del Rey JA, Dewitt TG, et al. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Arch Pediatr Adolesc Med*. 2007;161(1):44–49.
- 10 Higgins RSD, Bridges J, Burke JM, O'Donnell MA, Cohen NM, Wilkes SB. Implementing the ACGME general competencies in a cardiothoracic surgery residency program using 360-degree feedback. *Ann Thorac Surg*. 2004;77(1):12–17.
- 11 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(5):458–463.
- 12 Massagli TL, Carline JD. Reliability of a 360-degree evaluation to assess resident competence. *Am J Phys Med Rehabil*. 2007;86(10):845–852.
- 13 Musick DW, McDowell SM, Clark N, Salcido R. Pilot study of a 360-degree assessment instrument for physical medicine & rehabilitation residency programs. *Am J Phys Med Rehabil*. 2003;82(5):394–402.
- 14 Wood J, Collins J, Burnside ES, Albanese MA, Propeck PA, Kelcz F, 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;8(11):931–939.
- 15 Stark R, Korenstein D, Karani R. Impact of a 360-degree professionalism assessment on faculty comfort and skills in feedback delivery. *J Gen Intern Med*. 2008;23(7):969–972.
- 16 Hesketh EA, Anderson F, Bagnall GM, Driver CP, Johnston DA, Marshall D, et al. Using a 360 degrees diagnostic screening tool to provide an evidence trail of junior doctor performance throughout their first postgraduate year. *Med Teach*. 2005;27(3):219–233.
- 17 Greene JC, Caracelli VJ, Graham WF. Toward a conceptual framework for mixed-method evaluation designs. *Educ Eval Policy Anal*. 1989;11(3):255–274.
- 18 Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ*. 2000;321(7262):694–696.
- 19 Nasca TJ, Philibert I. Communities of practice and learning: disseminating their work. *J Grad Med Educ*. 2009;1(1):164–165.