Can 360-Degree Reviews Help Surgeons? Evaluation of Multisource Feedback for Surgeons in a Multi-Institutional Quality Improvement Project

Suliat M Nurudeen, MD, MPH, Gifty Kwakye, MD, MPH, William R Berry, MD, MPH, MPP, FACS, Elliot L Chaikof, MD, PhD, FACS, Keith D Lillemoe, MD, FACS, Frederick Millham, MD, MBA, FACS, Marc Rubin, MD, FACS, Steven Schwaitzberg, MD, FACS, Robert C Shamberger, MD, FACS, Michael J Zinner, MD, FACS, Luke Sato, MD, Stuart Lipsitz, ScD, Atul A Gawande, MD, MPH, FACS, Alex B Haynes, MD, MPH

BACKGROUND: Medical organizations have increased interest in identifying and improving behaviors that threaten team performance and patient safety. Three hundred and sixty degree evaluations of surgeons were performed at 8 academically affiliated hospitals with a common Code of Excellence. We evaluate participant perceptions and make recommendations for future use.

STUDY DESIGN: Three hundred and eighty-five surgeons in a variety of specialties underwent 360-degree evaluations, with a median of 29 reviewers each (interquartile range 23 to 36). Beginning 6 months after evaluation, surgeons, department heads, and reviewers completed follow-up surveys evaluating accuracy of feedback, willingness to participate in repeat evaluations, and behavior change.

RESULTS: Survey response rate was 31% for surgeons (118 of 385), 59% for department heads (10 of 17), and 36% for reviewers (1,042 of 2,928). Eighty-seven percent of surgeons (95% CI, 75%-94%) agreed that reviewers provided accurate feedback. Similarly, 80% of department heads believed the feedback accurately reflected performance of surgeons within their department. Sixty percent of surgeon respondents (95% CI, 49%-75%) reported making changes to their practice based on feedback received. Seventy percent of reviewers (95% CI, 69%-74%) believed the evaluation process was valuable, with 82% (95% CI, 79%-84%) willing to participate in future 360-degree reviews. Thirty-two percent of reviewers (95% CI, 29%-35%) reported perceiving behavior change in surgeons.

CONCLUSIONS: Three hundred and sixty degree evaluations can provide a practical, systematic, and subjectively accurate assessment of surgeon performance without undue reviewer burden. The process was found to result in beneficial behavior change, according to surgeons and their coworkers. (J Am Coll Surg 2015;■:1−8. © 2015 by the American College of Surgeons)

With increasing focus on a team-based approach to medicine, the role of the doctor is evolving into a member, and sometimes leader, of a multispecialty patient-centered health care team. To provide effective care within this environment, clinicians must possess a set of skills beyond technical ability. A growing body of

Disclosure Information: This study was supported by a grant from The Risk Management Foundation of the Harvard Medical Institutions, Inc.

Disclosures outside the scope of this work: Dr Berry is a paid consultant to The Risk Management Foundation of the Harvard Medical Institutions.

Support: The study design, analysis, and decision to publish were considered independently of these funding sources and individual data were not available to the insurer. The data collected by the institution will not be used for any purpose other than those stated in this article.

Abstract presented at the American College of Surgeons 100th Annual Clinical Congress, San Francisco, CA, October 2014.

Received April 30, 2015; Revised June 24, 2015; Accepted June 25, 2015.

From Ariadne Labs (Nurudeen, Kwakye, Berry, Lipsitz, Gawande, Haynes), Brigham and Women's Hospital (Nurudeen, Kwakye, Berry, Zinner, Lipsitz, Gawande), Harvard School of Public Health (Nurudeen, Berry, Lipsitz, Gawande, Haynes), Beth Israel Deaconess Medical Center (Chaikof), Massachusetts General Hospital (Lillemoe, Haynes), Boston Children's Hospital (Shamberger), Boston, South Shore Hospital, South Weymouth (Millham), North Shore Medical Center, Salem (Rubin), Cambridge Health Alliance (Schwaitzberg), and Risk Management Foundation of the Harvard Medical Institutions, Inc (Sato), Cambridge, MA.

Correspondence address: Alex B Haynes, MD, MPH, Massachusetts General Hospital, 55 Fruit St, Yawkey 7B, Boston, MA 02114. email: abhaynes@mgh.harvard.edu

http://dx.doi.org/10.1016/j.jamcollsurg.2015.06.017
ISSN 1072-7515/15

literature has identified the importance of nontechnical skills, such as communication and interpersonal behavior, in addition to technical competency.[1,2] Evaluations using simulated cases and objective structured clinical examinations have confirmed the value of these competencies, as well as provided methods for obtaining information on individual practitioners' strengths and weaknesses.[3] However, simulation can be costly and time consuming, and has other attributes that limit its use as an evaluation and improvement tool.[4] Clinical performance and competency have often been measured through a combination of performance evaluations and standardized testing. However, performance evaluations from a single source, such as a supervisor, subordinate, or patient, can have inherent inaccuracies, including inflated ratings, leniency, and the "halo and horn" effect.[5,6] Although standardized testing is a useful means of ensuring minimum proficiency in medical knowledge and clinical reasoning, it does not capture a complete sense of competency combining the technical and nontechnical aspects of care, and does not provide a reliable tool to distinguish exceptional physicians from marginally competent ones. [7]

Multisource feedback (MSF) has been a mainstay in performance evaluations in many industries for decades. This approach gathers feedback from multiple people occupying varying roles in an individual's work environment and serves to generate a comprehensive perspective on performance. More recently, hospital systems have used MSF as a way to measure physician performance.[8] Multisource feedback, often referred to as 360-degree feedback, has been incorporated into the recertification process in several countries.[9] By soliciting feedback from multiple sources within a physician's work environment, including peers, superiors, and subordinates, a more global assessment of performance is obtained, minimizing bias, including that based on race, sex, and age.[10] The result of the information received has been used as a method to guide professional development and to track employee progress over time.[11,12]

In this study, we assessed the value of an MSF program sponsored by a malpractice insurance company for a group of 8 diverse hospitals affiliated with a common university system. The program was deployed as part of a long-standing, surgical chief-led patient safety and quality collaborative. The collaborative had previously constructed a Code of Excellence (COE), an explicit description of behaviors expected of all surgeons within their departments. The 360-degree evaluation process was designed to assess progress toward these standards.

Earlier studies found a beneficial role for 360-degree feedback in various physician specialties.[13-16] However, effectiveness of MSF has been recognized to depend critically on how the program is implemented, how the feedback is given to subjects, and how institutional officials use the information.[17] We sought to describe the program deployed and determine through surveys the subjective accuracy of surgeon performance assessment and the effect of feedback on subsequent behavior.

METHODS

Setting

In 2005, The Risk Management Foundation of the Harvard Medical Institutions, Inc., the malpractice insurance and patient safety company insuring the Harvard-affiliated hospitals, convened a surgical safety and quality collaborative led by the surgical department heads across multiple institutions. This group has produced and published a number of system-wide improvement initiatives.[18-21] In 2011, the collaborative developed a COE defining a minimum standard of conduct expected of all affiliated surgeons in the following themes: service, respect, teamwork, excellence, ethical discipline, personal responsibility to patients, openness, education, humility, health, and conflict of interest. After its development, each department used a variety of methods to implement the code, ranging from formal presentations to signed endorsement by individual surgeons acknowledging their intent to behave in accordance with these standards.

The 360-degree evaluation process

From 2012 to 2013, the 8 participating hospitals implemented a 360-degree review process using a proprietary web-based system (PULSE 360 Program). The 360-degree tool consists of 40 questions that provide an assessment of an individual's professionalism, communication skills, interpersonal style, leadership, and teamwork approaches. A working group of surgeons revised and expanded the tool to fully capture the themes of performance covered by the COE. Each question was scored on a 5-item Likert scale based on level of agreement and was mapped to a COE theme. The theme focused on the physician's personal health was excluded from the 360 degree evaluation. A global COE score was created by taking the mean from all 10 COE themes and multiplying by 20, creating a scale from 20 to 100. Physicians were deemed as having concerning behavior if they scored <2 SDs below the mean in their COE global score. Participating institutions included 4 university-affiliated community hospitals and 4 academic medical centers. Participants included 385 attending surgeons from the departments of cardiac, thoracic, vascular, orthopaedic, plastic, and general surgery. Participation by surgeons was mandatory, with the stipulation that the results would be

used for no other purpose than physician development. Each surgeon was given the opportunity to select 20 to 30 individuals as evaluators, including peers, referring physicians, trainees, nurses, ancillary operating room staff, administrative assistants, and/or supervisors. The final list of reviewers was determined by the associated department or division head. This partial self-selection process has been demonstrated to improve perceptions of fairness and rater credibility by participants.[22] The vendor then emailed surveys and automated reminders to prompt reviewers to complete the 360-degree evaluation through an online portal. The review process occurred during a 2- to 3-month period, with staggered evaluation delivery to reduce reviewer fatigue. The number of reviews requested of an individual were also tracked and limited to 6 per reviewer. The reviews were deidentified and compiled into a report distributed to the attending surgeons. All evaluation results were anonymous and without distinguishing characteristics to prevent identification of the reviewers. Each department head determined how the reviews were distributed to surgeons and whether formal debriefing and/or follow-up coaching was provided.

Survey process

Beginning at 6 months after completion of each institution's 360-degree evaluation process, a web-based survey was distributed through the vendor's secure server to department heads, reviewers, and surgeons. The survey instruments were constructed by the study team with the assistance of a panel of subject-matter experts and customized for each recipient group. Each survey consisted of 8 to 10 multiple-choice and open-ended questions about the usefulness, accuracy, and sustainability of the 360-degree review process. The questions were scored using a 5-item Likert scoring system. The vendor deidentified the survey results and provided them to our research group for analysis.

Table 1. Mean Scores by Code of Excellence Theme

COE theme	Mean (95% CI)	Correlation to COE (95% CI)
Service	4.33 (4.29-4.37)	0.883 (0.848-0.911)
Respect	4.42 (4.37-4.47)	0.771 (0.595-0.884)
Teamwork	4.37 (4.30-4.44)	0.706 (0.457-0.863)
Excellence	4.46 (4.44-4.48)	0.690 (0.610-0.758)
Ethical discipline	4.42 (4.39-4.44)	0.665 (0.604-0.719)
Personal responsibility	4.38 (4.35-4.41)	0.686 (0.615-0.746)
Openness	4.34 (4.30-4.38)	0.900 (0.878-0.918)
Education	4.42 (4.36-4.47)	0.683 (0.471-0.829)
Humility	4.43 (4.38-4.48)	0.836 (0.761-0.892)
Conflict of interest	4.26 (4.23-4.28)	0.534 (0.444-0.615)

COE, Code of Excellence.

Provider characteristics and study results are presented using percentages for categorical variables and medians with interquartile ranges for continuous variables. For dichotomous variables, 2-sided 95% exact binomial CIs were used to calculate CIs for the unknown population percentage. To minimize the bias due to missing data when estimating the sample means and correlations among the COE themes, we performed multiple imputation, a statistically valid approach to use with missing data.[23] All analyses were conducted using Stata/IC 13 software (Stata Corp).

RESULTS

The 360-degree evaluation results

A total of 385 surgeons across 8 facilities participated in 360-degree evaluations. A composite score incorporating all themes of the COE equally was created for each participant. Mean COE score for all surgeons was 87.6 (95% CI, 86.9-88.3) on a scale of 20 to 100. Mean score by each COE theme is displayed in Table 1, as well as their correlation to the global score. The themes of service, openness, and humility seemed to most correlate with a physician's overall score. The items least correlated with the overall score were in the areas of excellence, ethical discipline, personal responsibility, education, and conflict of interest. A total of 21 surgeons (5%) were classified as having overall concerning behavior based on their composite score (mean 76.4; 95% CI, 74.6-78.3).

Study participants

Participants in the follow-up study consisted of individuals from the following groups: department heads, surgeons, and reviewers. Demographic data were requested for the surgeons and their reviewers (Tables 2 and 3). Reviewers were primarily composed of physician peers (28.9%), including fellow surgeons, referring physicians, and anesthesiologists; administrative staff (19%); and nursing staff (20%). Forty percent of surgeons reported having more than 16 years of professional experience. One third of surgeons reported being at their particular institution for more than 16 years. The reviewers had a similar distribution of years of experience and institutional longevity.

A total of 2,928 reviewers completed 360-degree evaluations for at least 1 attending surgeon, with a median of 29 reviewers (interquartile range 23 to 36) per surgeon. One thousand and forty-two individuals responded to the survey, yielding a response rate of 36%. Of these, 96 did not recall completing a 360-degree evaluation and were excluded from the final analysis. In addition, 10 of 17 department or division heads (59%) and 118 of

Table 2. Demographic Data: Surgeon Respondents

Surgeon characteristics	n	%
Department		
Cardiac	3	5.0
General	19	31.7
Orthopaedic	13	21.7
Other	7	11.7
Declined to answer	18	30.0
Years of experience in specialty		
1 to 5	7	11.7
6 to 10	6	10.0
11 to 15	11	18.3
16+	24	40.0
Declined to answer	12	20.0
Years of experience at current hospital		
1 to 5	13	21.7
6 to 10	10	16.7
11 to 15	7	11.7
16+	18	30.0
Declined to answer	12	20.0

385 surgeon participants (31%) responded to the survey. A portion of the surgeon participants were asked to complete a reviewer survey, as they had performed evaluations of their peers.

Table 3. Demographic Data: Reviewer Respondents

Reviewer characteristics	n	%
Reviewer role		
Administrative staff	192	19.1
Ancillary staff	46	4.6
Nurse practitioner/physician assistant	63	6.3
General nursing staff	198	19.7
Physician peer	290	28.9
Other	12	1.2
Declined to answer	203	20.2
Years of experience in specialty		
0	3	0.3
1 to 5	147	14.6
6 to 10	184	18.3
11 to 15	161	16.0
16+	404	40.2
Declined to answer	105	10.5
Years of experience at current hospital		
0	2	0.2
1 to 5	149	14.8
6 to 10	223	22.2
11 to 15	183	18.2
16+	342	34.1
Declined to answer	105	10.5

Behavior change

Sixty-three percent (95% CI, 49%-79%) of participants reported making changes to their practices based on the results of their 360-degree evaluation. Sixty percent (95% CI, 26%-88%) of department heads noted an overall improvement in their staff's behavior, especially in the areas of communication and professionalism. Thirty-two percent (95% CI, 29%-35%) of reviewers reported an appreciable change in staff behavior after 360-degree evaluation completion (Fig. 1).

Accuracy of feedback

Surgeon participants were surveyed on their perceptions of the accuracy of their reports. Eighty-seven percent (95% CI, 75%-94%) believed that the information received from external raters was accurate (Fig. 2). Seventy-seven percent (95% CI, 64%-87%) of surgeon participants were still willing to participate in a repeat 360-degree evaluation (Fig. 3). Eighty percent (95% CI, 44%-97%) of department heads thought the reports correctly demonstrated the performance of the surgeons in their departments (Fig. 2). Sixty percent (95% CI, 26%-88%) of department heads reported that the 360-degree evaluations confirmed the identity of previously concerning surgeons within their departments, and 20% (95% CI, 3%-56%) reported that they helped to identify surgeons without a previous record of concerning behavior.

Willingness for future participation

Eighty percent (95% CI 44%-97%) of department heads and 85% (95% CI, 83%-87%) of reviewers reported that they would be willing to participate in the evaluation process again (Fig. 3). Barriers to participating in the future included concerns for "survey fatigue," concerns about inadequate benefit, and the time investment required of surgical staff (Table 4).

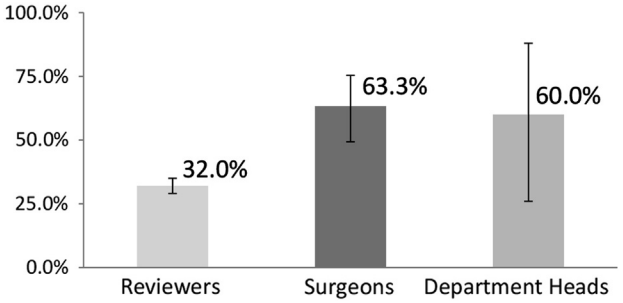

Figure 1. Perceived or enacted degree of behavior change. Sixty percent of surgeons (95% CI, 49%-75%) reported making changes to their practice based on this feedback. Thirty-two percent of reviewers (95% CI, 29%-35%) reported perceiving change in reviewed surgeons.

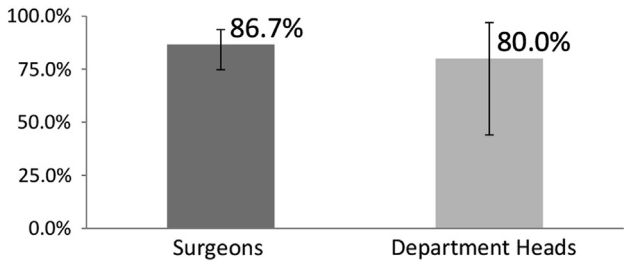

Figure 2. Perceptions regarding accuracy of 360-degree feedback. Eighty-seven percent of surgeon (95% CI, 75%-94%) agreed that reviewers had provided accurate feedback, and 8 of 10 department heads believed the feedback accurately reflected the performance of surgeons within their departments.

In addition, 5% of reviewers (95% CI, 4%-7%) reported experiencing some form of negative repercussions due to their participation.

DISCUSSION

We found that our collaborative of surgical leaders from 8 hospitals was able to implement a web-based 360-degree evaluation program for a large number of surgeons without undue burden on them or on reviewers. In addition, we found that this program was well supported by the majority of surgeons and reviewers in our study. Most of the respondents reported making practice changes as a consequence of the feedback, and one third of the reviewers observed visible improvements in professional behavior and practice in line with the COE the surgical leaders articulated. As the complexity of medical care increases, so do the responsibilities for teams of interlinking disciplines to achieve successful outcomes and patient-centered care. The surgical leaders who responded believed that this approach of using MSF was a valuable adjunct to existing mechanisms of evaluation and feedback at their disposal and intended to continue with the program.

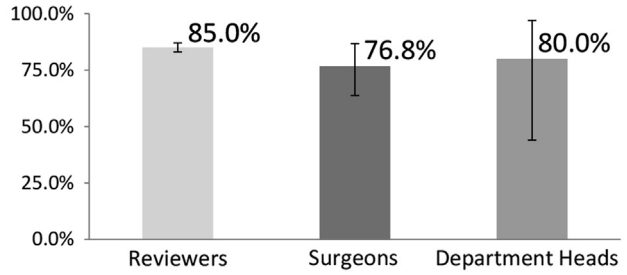

Figure 3. Willingness to participate in future participation. Seventy percent of reviewers (95% CI, 69%-74%) believed the process in general was valuable, with 85% (95% CI, 83%-87%) willing to participate in future 360-degree evaluations. Similarly, 76.8% (95% CI, 64%-87%) of surgeons and 80% of department heads were interested in future participation.

Table 4. Factors Affecting Willingness of Reviewers to Participate in Future 360-Degree Evaluations

Factors affecting willingness to participate	%	95% CI
No barriers	40.8	37.3-43.4
Fatigue with completing evaluations	23.3	20.4-25.7
Concern for time investment it requires	23.2	20.3-25.6
Perception of inadequate benefit	20.7	18.0-23.2
Concern of experiencing negative repercussions	12.2	10.1-14.2

Respondents had the ability to choose multiple options.

Multisource feedback is not a new concept, and has been a mainstay in other industries for decades. Some of the earliest work in MSF from the Center for Creative Leadership emphasized the importance of feedback in professional development and described the feedback-poor environments of most organizations. Since that time, it has been estimated that as many as 90% of Fortune 500 companies use MSF in some form.[24] The frequency in which these types of evaluations are performed is dependent on the resources of the organization and the intent of the 360-degree review process. A study by Walker and colleagues[25] describes a 5-year upward feedback program in which 252 managers received feedback ratings from their direct reports at multiple time points. Those managers that initially rated poorly in comparison with their peers demonstrated incrementally improved ratings during the trial period.

How MSF programs are implemented makes a significant difference in outcomes. Bracken[26] described 4 major design components to elicit behavior change using MSF: relevant content, credible data, accountability, and organization-wide participation. Relevant content was addressed in this program by mapping all questions to a category within the previously established COE, well known to all surgeon participants. This study used a customized version of a validated instrument provided by the vendor. Previous MSF research has demonstrated the efficacy of standardized survey tools, such as the UK's Sheffield Peer Review Assessment Tool[27] and the Physician Achievement Review[28] used by The Netherlands and Canada. There is no consensus on use of a standardized evaluation vs a customized tool. The generalizability of this study might be limited due to the institution-specific tool used. The credibility of the data was improved by ensuring that physicians were rated by more than 10 individuals from varying aspects of their work environment. It was also enhanced by the surgeon's

ability to self-select a portion of their evaluators. Allowing department and division heads to approve the final list of evaluators was designed to reduce the risk of overly positive ratings secondary to rater familiarity. Surgeons were held accountable for their results by having an in-person review with either their department heads or an external coach. The receipt of a comprehensive report from multiple parties serves to reinforce the feedback received from evaluations. Enlisting an outside party, such as a supervisor or coach, to deliver evaluation results assists with participant receptiveness.[26] Due to logistical concerns, the mechanism of feedback delivery was variable across institutions, creating a challenge for determining best practices.

Organization-wide participation was encouraged by first including it as part of a multi-institutional safety collaborative guided by the surgical leaders, and by having them agree to be the first evaluated before adoption by all surgeons within the department. This has been demonstrated previously as a method to facilitate widespread implementation and improve feedback acceptance.[6] Participating surgeons did express reservations about the potential use of the feedback in a punitive manner. Before beginning the process, surgeons were ensured the results would be for the purpose of physician development only. One concern of the 360-degree feedback process is that it relies on the ratee's ability to process the information provided and use it to change behaviors found to be of concern. It is this variability in a ratee's capacity to process feedback that can be most limiting with this assessment tool. Lipsett and colleagues[29] described an inability of low-performing surgical residents to identify their weaknesses as perceived on a global evaluation. This has previously been attributed to deficits in individual insight and cognition, as described by Kruger and Dunning.[30] In this study, a discrepancy was noted between surgeon-reported behavior change and those perceived by reviewers. It is possible that the difference seen is a result of the varying aspects of a clinician's practice. Although a physician might make changes in one environment, those changes might not be perceived in another area of their work flow. For example, a surgeon's efforts to ensure that he or she arrives to the operating room in a timely manner would not be visible to the administrative staff in their office. It is also likely that although changes might seem large to an individual, they might not necessarily be perceived that way by external parties.

For the entire group of surgeons, the COE themes that best correlated with their scores were associated with service, openness, and humility. This is contrasted by the fact that certain aspects of the COE did not align with the global score in all surgeons. These incongruent areas included education, excellence, ethical discipline, personal responsibility, and conflict of interest. It might be suspected that the latter 3 themes are difficult for individuals to interpret, dependent on their area of interaction with the clinician, as these were also less correlated in the entire group. What is most interesting is that the highly technically based theme, excellence, as well as perceived medical knowledge, did not correlate with overall perceptions of physician behavior. Traditionally, surgeon evaluation has been based largely on technical proficiency and clinical knowledge, but it is clear from these data that coworker discernment of performance encompasses more than these aspects alone.[31] Many nonmedical organizations have enlisted the assistance of "feedback coaches" to help individuals undergoing 360-degree reviews to interpret their final evaluation and coach them in skills necessary to modify these behaviors.[32] These individuals can serve an interdisciplinary role, keeping in mind the needs of the individual as well as the organization.[33] The use of formalized debriefing and additional coaching after initial results were distributed was variable across institutions in this study, however, due to both expense and logistics. A more uniform practice of interval coaching and training of potential coaches might have resulted in more pronounced behavior alteration and improved the chances of these alterations persisting over time. The interval reinforcement of preferred behaviors could result in a more sustainable change over time.

Limitations

One might cite the affiliation with a single institution as a limitation to the study. However, although the participating facilities are indeed affiliated with a single medical school, they represent a wide range of hospitals, including academic medical centers and university-affiliated community hospitals with a voluntary medical staff model. These facilities treat a variety of disease processes, as well as patients from pediatrics to the geriatric population. A limitation of the study is that the response rate, which is acceptable, is lower than ideal. Conclusions can be drawn from the results observed, but the opinions of nonresponders cannot be fully taken into account. One must also note the self-selection bias of those that did choose to participate in the follow-up survey. Individuals that chose not to participate in the follow-up survey might have been

less enthusiastic about the process and therefore not inclined to complete the survey. Individual physician ratings were skewed toward a more favorable range. This is often the case when using Likert scoring for such evaluations. Despite this positive skew, the formation of a normative scale allows for the identification of outliers whose behaviors are deemed less satisfactory compared with their peers. Despite these limitations, to experience the amount of change demonstrated in the setting of variable coaching practices across institutions is telling. In any multirater feedback process, the true measure of success is in whether the information gathered motivates an individual to change, as was seen in this study from the perspective of those that were evaluated, as well as those that evaluated others. Ideally, future research efforts would link physician behavior to defined end points, such as malpractice claim rates and clinical outcomes, to reinforce the role environment plays in patient safety.

CONCLUSIONS

Overall, we found that the process was deemed accurate by multiple parties. Despite the expressed concerns of rater fatigue and time investment required, the majority of surgeon participants, reviewers, and department heads would be interested in future participation. Only a small percentage of reviewers reported experiencing negative repercussions, although theoretically this should be nonexistent. The feedback received by individuals was relevant enough to induce change noticeable by their colleagues. This satisfies the true goal of this process, which is to elicit an individual to look introspectively and adapt, especially in those individuals deemed lacking in this area of competence. Based on these findings, a comprehensive 360-degree evaluation program in conjunction with a departmental commitment to quality improvement is an effective means of assessing surgeon nontechnical and interpersonal skills and serves a role in behavior modification.

Author Contributions

Study conception and design: Nurudeen, Kwakye, Berry, Chaikof, Lillemoe, Millham, Rubin, Schwaitzberg, Shamberger, Zinner, Sato, Lipsitz, Gawande, Haynes

Acquisition of data: Nurudeen, Kwakye, Berry, Gawande, Haynes

Analysis and interpretation of data: Nurudeen, Kwakye, Berry, Lipsitz, Gawande, Haynes

Drafting of manuscript: Nurudeen, Kwakye, Berry, Chaikof, Lillemoe, Millham, Rubin, Schwaitzberg, Shamberger, Zinner, Sato, Lipsitz, Gawande, Haynes

Critical revision: Nurudeen, Kwakye, Berry, Chaikof, Lillemoe, Millham, Rubin, Schwaitzberg, Shamberger, Zinner, Sato, Lipsitz, Gawande, Haynes

Acknowledgment: This study was supported by a grant from The Risk Management Foundation of the Harvard Medical Institutions, Inc. The study design, analysis, and decision to publish were considered independently of these funding sources and individual data were not available to the insurer. The data collected by the institution will not be used for any purpose other than those stated in this manuscript. The authors would like to thank the Harvard Surgical Chiefs Collaborative and the Physician Development Program/PULSE 360 Program.

REFERENCES

1. Tumerman M, Carlson LM. Increasing medical team cohesion and leadership behaviors using a 360-degree evaluation process. WMJ 2012;111:33–37.
2. Flin RH, O'Connor P, Crichton M. Safety at the Sharp End: A Guide to Non-Technical Skills. Burlington, VT: Ashgate Publishing, Ltd.; 2008.
3. Harden RM, Gleeson F. Assessment of clinical competence using an objective structured clinical examination (OSCE). Med Educ 1979;13:39–54.
4. Wallace J, Rao R, Haslam R. Simulated patients and objective structured clinical examinations: review of their use in medical education. Adv Psychiatr Treat 2002;8:342–348.
5. Bettenhausen KL, Fedor DB. Peer and upward appraisals: a comparison of their benefits and problems. Group Organ Manage 1997;22:236–263.
6. Wood L, Hassell A, Whitehouse A, et al. A literature review of multi-source feedback systems within and without health services, leading to 10 tips for their successful design. Med Teach 2006;28:e185–e191.
7. Landon BE, Normand ST, Blumenthal D, Daley J. Physician clinical performance assessment: rospects and barriers. JAMA 2003;290:1183–1189.
8. Donnon T, Al Ansari A, Al Alawi S, Violato C. The reliability, validity, and feasibility of multisource feedback physician assessment: a systematic review. Acad Med 2014;89:511–516.
9. 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.
10. Drew G. A '360' degree view for individual leadership development. J Manage Dev 2009;28:581–592.
11. Brutus S, Fleenor JW, London M. Does 360-degree feedback work in different industries? A between-industry comparison of the reliability and validity of multi-source performance ratings. J Manage Dev 1998;17:177–190.
12. Sargeant J, Bruce D, Campbell CM. Practicing physicians' needs for assessment and feedback as part of professional development. J Contin Educ Health Prof 2013;33[Suppl 1]: S54–S62.
13. Bahaziq W, Crosby E. Physician professional behaviour affects outcomes: a framework for teaching professionalism during anesthesia residency. Can J Anaesth 2011;58:1039–1050.

14. Lockyer JM, Violato C, Fidler H. The assessment of emergency physicians by a regulatory authority. Acad Emerg Med 2006;13:1296–1303.
15. Lockyer JM, Violato C, Fidler H. A multi source feedback program for anesthesiologists. Can J Anaesth 2006;53:33–39.
16. 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.
17. Lockyer J. Multisource feedback in the assessment of physician competencies. J Contin Educ Health Prof 2003;23:4–12.
18. Greenberg CC, Diaz-Flores R, Lipsitz SR, et al. Bar-coding surgical sponges to improve safety: a randomized controlled trial. Ann Surg 2008;247:612–616.
19. Arriaga AF, Bader AM, Wong JM, et al. Simulation-based trial of surgical-crisis checklists. N Engl J Med 2013;368:246–253.
20. Arriaga AF, Gawande AA, Raemer DB, et al. Pilot testing of a model for insurer-driven, large-scale multicenter simulation training for operating room teams. Ann Surg 2014;259:403–410.
21. ElBardissi AW, Regenbogen SE, Greenberg CC, et al. Communication practices on 4 harvard surgical services: a surgical safety collaborative. Ann Surg 2009;250:861–865.
22. Becton JB, Schraeder M. Participant input into rater selection: potential effects on the quality and acceptance of ratings in the context of 360-degree feedback. Public Pers Manage 2004;33:23–32.
23. Rubin DB. Multiple Imputation for Nonresponse in Surveys, Vol 81. Hoboken, NJ: John Wiley & Sons; 2004.
24. Lepsinger R, Lucia AD. The Art and Science of 360 Degree Feedback. Hoboken, NJ: John Wiley & Sons; 2009.
25. Walker AG, Smither JW. A five-year study of upward feedback: what managers do with their results matters. Person Psychol 1999;52:393–423.
26. Bracken DW, Rose DS. When does 360-degree feedback create behavior change? and how would we know it when it does? J Bus Psychol 2011;26:183–192.
27. Archer J, Davies H. Sheffield Peer Review Assessment Tool for Consultants (SPRAT): Screening for Poorly Performing Doctors. Bern, Switzerland: Association of Medical Education of Europe; 2003.
28. Hall W, Violato C, Lewkonia R, et al. Assessment of physician performance in Alberta: the physician achievement review. CMAJ 1999;161:52–57.
29. Lipsett PA, Harris I, Downing S. Resident self-other assessor agreement: influence of assessor, competency, and performance level. Arch Surg 2011;146:901–906.
30. 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.
31. Arora S, Sevdalis N, Suliman I, et al. What makes a competent surgeon? Experts' and trainees' perceptions of the roles of a surgeon. Am J Surg 2009;198:726–732.
32. Seifert CF, Yukl G, McDonald RA. Effects of multisource feedback and a feedback facilitator on the influence behavior of managers toward subordinates. J Appl Psychol 2003;88:561–569.
33. Goodstone MS, Diamante T. Organizational use of therapeutic change: strengthening multisource feedback systems through interdisciplinary coaching. Consult Psychol J 1998;50:152.