

A randomized, controlled crossover study to discern the value of 360-degree versus traditional, faculty-only evaluation for performance improvement of anesthesiology residents.

Authors: Jeffrey S. Berger, M.D., M.B.A., Eric Pan, M.D., Jason Thomas, M.D.

Institution: The George Washington University Medical Center

Department of Anesthesiology and Critical Care Medicine

Corresponding author: Jeffrey S. Berger, M.D., M.B.A.

The George Washington University Medical Center

Department of Anesthesiology and Critical Care Medicine

900 23rd Street, NW Suite G-2092

Washington, DC 20037

Phone: 202-715-5296

Fax: 202-715-4759

Email: jberger@mfa.gwu.edu

Financial support: None

ABSTRACT

BACKGROUND

360-degree evaluation, or multi-rater feedback, is a means of providing evaluation from a variety of stakeholders. The Accreditation Council for Graduate Medical Education (ACGME) lists 360-degree feedback as a recommended method for evaluating residents. Our study determines if 360-degree evaluation, as compared to traditional evaluation, affords anesthesiology residents greater potential for performance improvement.

METHODS

After IRB exemption and resident consent to participate, sixteen anesthesiology residents of various training levels at The George Washington University Medical Center were randomly assigned to receive either 360-degree evaluation or traditional evaluation. Mid-way through the study, the groups were crossed over. Three faculty members blinded to which type of evaluation each resident received evaluated all residents pre-study, midway, and at the conclusion of the study according to the six core competencies set forth by the ACGME. The 360-degree study included evaluations by faculty (traditional), self, medical students, nursing staff and patients.

RESULTS

Performance improvement in all core competencies demonstrated a trend toward greater scoring for residents who received early exposure to 360-degree evaluation compared to later exposure. Paired t-Tests demonstrated significance for resident performance improvement with early 360-degree evaluation versus early traditional evaluation for the Interpersonal and Communication Skills core competency over the entire study. Systems-based Practice and Practice-Based Learning and Improvement suggested improvements, but failed to reach statistical significance ($P = 0.09, 0.07$ respectively).

WORDS: 220

KEYWORDS

360-degree evaluation; Evaluation; Feedback; Multi-rater feedback; Core competencies; Resident education

SUMMARY

Does a 360-degree evaluation, as compared to traditional, faculty-only evaluation afford anesthesiology residents greater potential for performance improvement? Sixteen residents were exposed to a prospective, single-blinded, crossover design study to determine the answer. For 360-degree evaluations, residents were rated according to the ACGME core competencies after adjusting to Program Director review of evaluations from self, faculty, nursing staff, medical students and patients.

A randomized, controlled crossover study to discern the value of 360-degree versus traditional, faculty-only evaluation for performance improvement of anesthesiology residents.

INTRODUCTION

In the business world, the 360-degree evaluation, also known as multi-rater or multi-source feedback, is a means of providing evaluation from a variety of stakeholders. Input comes from subordinates, peers, managers, and in some cases customers, rather than the traditional supervisor-subordinate-only assessment model. The idea is that an individual can gain a broader perspective of vertically (boss, customer) and horizontally (colleague) integrated perceptions, increase awareness of competencies, and obtain a better understanding of critical performance aspects.¹ Originally used in the military and in the business world to train high-level managers, by the late 1990s human resource departments had widely instituted the 360-degree evaluation model throughout their workforces. Following suit, residency training programs have begun to adopt this comprehensive assessment tool for the purpose of resident evaluations, however evidence of benefit is lacking.²

The Accreditation Council for Graduate Medical Education (ACGME) is a private, non-profit council that evaluates and accredits medical residency programs in the U.S. In 2007, the ACGME established six core competencies that all programs must integrate into their curricula. In addition to addressing the six core competencies, the ACGME requires evaluative instruments to provide valid, reliable data that is feasible to obtain and externally valid.³ The ACGME dictates that evaluation of residents must yield valuable information which we interpret as a mandate for performance improvement. The 360-degree evaluation is one of 13 methods recommended by the ACGME. The ACGME's *Toolbox of Assessment Methods*, published in 2000, states that there are no published reports of the effectiveness of the use of the 360-degree evaluation in graduate medical education.⁴

Anesthesiology residency programs, perhaps more than other specialties, provide an excellent platform to study evaluation tools due to the discrete nature of interactions between anesthesiology resident and potential raters. Further, the brief and intense nature of the working environment maximizes the potential for performance improvement with adequate feedback. While many other specialties interact with nursing staff, colleagues, and patients indirectly via written or typed “orders,” most anesthesiologists must personally manage team dynamics and patient satisfaction. This personal, discrete, brief, and intense practice repeats with enough frequency to potentially measure performance improvement with a 360-degree evaluation tool. We aim to study whether 360-degree evaluation feedback enhances the performance of anesthesiology residents when compared to the standard faculty-only evaluations. We used a randomized, controlled, crossover study to refute the null hypothesis that there is no difference in resident performance following a 360-degree versus a traditional evaluation.

METHODS

Following Institutional Review Board exemption, we divided sixteen residents into two groups using randomization software, repeating the randomize command until an equal number of residents were in each group. The two groups, group “1st 360” and group “1st Traditional” represented all three Clinical Anesthesia years according to Table 1. Residents received an information sheet detailing the aims of the study and consented to participate accordingly. Residents were not blinded to group designation and intention to treat analysis was employed such that residents remained in the group they were initially randomized to regardless of the number of evaluations obtained for a particular resident.

During the first four weeks of the study, group “1st 360” received 360-degree evaluations and group “1st Traditional” received faculty-only evaluations. Residents receiving traditional evaluations obtained feedback from faculty and they conducted a self-evaluation in accordance with the current department

policies. Faculty, medical students, patients, and nurses evaluated the 360-degree evaluation group of residents in addition to completing self evaluations. After four weeks, the groups were switched such that 1st 360 only received faculty feedback and 1st Traditional received 360-degree feedback on performance. The study concluded at the end of eight weeks.

Evaluations were collected during the first two weeks of each month. The Evaluation Questionnaires were distributed in one of three ways: first, the self-evaluation was given to the residents via e-mail attachment; second, the faculty completed evaluations on the computer-based, *E-value* (Advanced Informatics, LLC, Minneapolis, MN) system; third, the medical students, patients and nursing staff completed evaluations that were distributed and collected directly by the investigators at assigned times. The questionnaires were one page in length and permitted written comments in addition to five-point, Likert-scaled responses to competency-based evaluation questions (Fig. 1). Evaluators, blinded to groupings, were asked to evaluate all residents throughout the study.

Period one was defined as Time = zero to four weeks. Period two was Time = four weeks to eight weeks; period three was Time = zero to eight weeks (Table 1). The timing for evaluation collections was as follows: 1) faculty evaluations were collected continuously with frequent reminders; 2) self-evaluations were completed and reviewed with the resident's advisor and program director prior to Time = two weeks; medical students provided evaluations every two weeks; 3) investigators distributed and collected evaluations from available nurses and patients daily during the collection periods. Pictures of residents were available for evaluators.

At Time = two weeks, groups 1st 360 and 1st Traditional received personalized feedback from the Program Director. Then again at week six, the two resident groups were given feedback from the Program Director, however according to the cross-over protocol, group 1st 360 received faculty-only feedback while

1st Traditional received the full, 360-degree evaluation. As Table 1 suggests, feedback was based solely on evaluations collected in the immediate two weeks prior to the meeting with the Program Director.

Meetings consisted of an individual discussion with the Program Director. During these meetings, the Program Director would review summative evaluation data from the E-value system for faculty evaluations, including all faculty comments. For residents receiving 360-degree feedback, evaluations from students, nursing staff and patients were reviewed with the residents and all comments were read together. Following a review of all evaluations, the resident and Program Director had a roughly 5-minute intervention to address any deficiencies noted in the evaluations and formulate a plan for improvement.

At Time = zero, four and eight weeks three “control” faculty evaluators, blinded to the evaluation group of any particular resident, evaluated all sixteen residents according to a core-competency-based, Likert scale (Fig. 2). They were instructed to mark an “X” along the line corresponding to a particular competency, from zero to one hundred. The “X” was measured in millimeters from the left, or zero position. The control evaluators were selected based on their long-standing experience in evaluating residents and their detailed knowledge of the core competencies. Additionally, control evaluators all directed the operating rooms one to two times per week and served as clinical anesthesiologists in a wide array of cases; hence control evaluators had extensive resident exposure on a daily basis. Residents were blinded as to which faculty were the control evaluators. Residents were evaluated during rotations at the following sites: The George Washington University Hospital’s main Operating Suite and the Labor and Delivery Unit. Eligible rotations for this study included any that allowed for control evaluators to supervise and observe the residents during the post-feedback portions of the study (T = two to four weeks and six to eight weeks). Rotations included: General, Advanced Clinical, Obstetric, Neuro, Cardiothoracic, and Regional Anesthesiology. Residents review core competency based goals and objectives prior to commencing each

rotation via the Departmental intranet; consequently, core competency improvement should be anticipated with each of the rotations included above.

SAS, Version 9.1 was used for statistical analysis. For each of the three time periods (zero to four, four to eight, and zero to eight weeks), the change in “control” evaluators’ scores of residents in the 360-degree and traditional feedback groups were calculated and between group comparisons made. P-values of ≤ 0.05 were considered significant. Goodness-of-Fit Testing confirmed that a T-Test should be employed as a modified normal distribution. Paired T-tests analyzed the averages for improving – or worsening – scores of the sample population.

RESULTS

Subject characteristics were noted on Table 2. All three residency classes were represented in equal numbers. Table 2 compared average test scores, sex, and number of evaluations of each type for both the 1st 360 and the 1st Traditional groupings.

The results of the Student’s T-Test analysis were summarized in Table 3. Periods one and two revealed no statistically significant differences in resident performance in the group that received 360-degree feedback versus the group that received traditional feedback. The Systems-Based Practice core competency approached statistical significance for period one ($P = 0.09$).

Over the duration of the study, period three, 1st 360 residents scored higher in all core competencies than residents who received 360-degree feedback in the second half of the study (1st Traditional). Additionally in period three, the residents improved their performance on Interpersonal and Communication Skills to a statistically significant degree with early exposure to 360-degree feedback ($P = 0.05$). Rater comments mainly referenced the Interpersonal and Communications Skills competency. Resident performance improvement with Practice-Based Learning and Improvement almost reached significance over the same period with early 360-degree feedback ($P = 0.07$).

DISCUSSION

Resident performance with respect to Interpersonal and Communication Skills significantly improved after early exposure to 360-degree evaluation. With respect to the core competency of Interpersonal and Communication Skills, it seems intuitive that the 360-degree evaluation model would yield greater improvement than traditional feedback. In this study, hand-written, “additional” comments from nurses, patients, and medical students almost exclusively related to interpersonal skills and communication issues. By incorporating 360-degree feedback, residents can collaborate with patients, their families, and other health professionals to hone critical interpersonal and communication skills.

The 1st 360 group also demonstrated a trend towards higher scoring from the control evaluators in all core competencies over the course of the study, period three. It is possible that a “carry-over” effect was responsible for this finding. Residents who initially received 360-degree feedback modified their own practices and behaviors, within the context of the six core competencies, for the duration of study. One could speculate that programs might benefit from providing 360-degree feedback during a discrete period early in the academic year, avoiding a year-round effort and the resources needed. Alternatively, when considering the entire study as compared to discrete periods one or two, confounding variables may be more important in a crossover study. In particular, the higher In-Training Examination average (48% vs. 25%) for the 1st 360 group may explain the higher scores.

Data analysis from periods one and two revealed no statistically significant differences in resident performance in the group that received 360-degree feedback versus the group that received traditional feedback. This finding suggests that the 360-degree feedback model was not successful in improving resident performance with respect to the six core competencies outlined by the ACGME. Nevertheless, the power of the study may not be sufficient to detect a subtle improvement for those receiving the 360-degree evaluation over those receiving traditional feedback. This would represent type II error and could possibly

be corrected by enrolling a larger number of subjects. Improvement in scores rating competency in Systems-Based Practice approached significance during period one ($P=0.09$). If such a statistical trend holds in a larger sample, it is conceivable that a 360-degree feedback model may particularly increase resident awareness of and responsiveness to the Systems-Based Practice competency. In order to show an educationally important change in scores, we recommend a multi-institutional follow-up study with no fewer than two hundred residents enrolled.

The methodology we employed to quantitatively measure performance improvement with 360-degree evaluation had important limitations: the study period was too brief (eight weeks), mitigating the potential for steady state measurements. A follow-up study would need to increase the period of time that residents adjust behaviors based on the Program Director's feedback from two weeks to at least three months. Further, the crossover nature of the study design allowed residents to serve as their own control and thus minimized the influence of confounding variables (such as the significant discrepancy between ITE scores for the two groups). This efficiency "balanced" the study; however, type II error may still have occurred. Crossover study design is also subject to carry-over effects, as noted in the discussion above. The statistical method utilized to generate results likewise has some potential weaknesses: the use of multiple t-tests in this study introduced the possibility of type I error; and, the significance reported for results is only marginally conclusive.

Our investigation reviews the prospects for anesthesiology resident performance improvement using the 360-degree evaluation tool. A few studies investigating the utility of 360-degree evaluation have been attempted in various other medical specialties.^{2, 5-8} *Brinkman et. al.* concluded that 360-degree feedback had a positive effect on communication skills and professional behavior among 36 pediatric residents.⁵ *Joshi, Ling and Jaeger* studied the 360-degree evaluation and its effect on graduate medical education in the field of obstetrics and gynecology. Their group found 360-degree evaluation reliable and useful for assessment of resident interpersonal and communication skills.⁶ Physical medicine and

rehabilitation resident performance was evaluated in another study using a web-based, multi-source feedback model. The investigation revealed that senior residents were consistently assigned higher ratings compared to junior residents. Hence, the authors concluded that the 360-degree method was reliable, feasible and valid.² Despite such sentiments across the gamut of medical specialties, an article in *Surgical Concepts* in 2004 questioned the performance improvement potential of 360-degree evaluation while emphasizing the extra work required to conduct the multi-rater evaluation tool.⁷

Compared to receiving traditional feedback from faculty-only, residents improved their performance in Interpersonal and Communication Skills after first receiving 360-degree feedback. This method of feedback may also facilitate developing the competency of Practice-Based Learning and Improvement although educational studies with larger sample sizes are needed to confirm the observed trends.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Christina Pinkston for her assistance with statistical analysis for this study.

REFERENCES

1. Adley, T. Guide to 360 Degree Feedback Deployment, <http://www.visionmetrics.net/survey-resources/360-degree-feedback-software.aspx>, accessed Oct 7, 2008.
2. Carline, JD, Massagli, TL. Reliability of a 360-degree evaluation to assess resident competence. *Am J Phys Med & Rehab* 2007; 86(10): 845-852.
3. Accreditation Council for Graduate Medical Education Outcomes Project. <http://www.acgme.org/outcome/comp/compMin.asp>, accessed on Oct 10, 2008.
4. Accreditation Council for Graduate Medical Education. *Toolbox of Assessment Methods*, Version 1.1, 2000.

5. Brinkman, WB, Geraghty, SR, Lanphear, BP, Khoury, JC, Gonzalez del Rey, JA, Dewitt TG, Britto, MT. Effect of Multisource Feedback on Resident Communication Skills and Professionalism: A Randomized Controlled Trial. *Arch of Ped & Adol Med* 2007; 161(1): 44-9.
6. Joshi, R, Ling, FW, Jaeger, J. Assessment of a 360-degree instrument to evaluate residents' competency in interpersonal and communication skills. *Academic Med* 2004; 79(5): 458-463.
7. Weigelt, JA, Brasel, KJ, Bragg, D, Simpson, D. The 360-degree evaluation: Increased work with little return? *Current Surg* 2004; 61(6): 616-626.
8. Higgins, RS, 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. *An Thor Surg* 2004; 77(1): 12-17.

LEGEND

Figure 1. Sample 360-degree evaluation: medical student

Figure 2. Control faculty evaluation

Table 1. Study timeline

*Control faculty were blinded to resident groupings. Program Director and residents were blinded to which faculty were controls and did not review control faculty evaluations during the study period.

Table 2. Subject characteristics

* Average number per resident for Time = 0 – 2 weeks

** Average number per resident for Time = 4 – 6 weeks

‡ Range of number of evaluations completed for each resident

Table 3. Student's T-Test Results

“1st 360” is the grouping of residents who were exposed to the 360-degree evaluation for the first 4 weeks of the study. “1st Trad” is the grouping of residents who only received a traditional, faculty-only evaluation for the first four weeks of the study. “Mean Δ ” represents the change in the averaged competency score of the three control evaluators from beginning to end of the time period. P-values of ≤ 0.05 were recognized as significant.