Repaying in Kind: Examination of the Reciprocity Effect in Faculty and Resident Evaluations ☆

Aimee K. Gardner, PhD, and Daniel J. Scott, MD

Department of Surgery, University of Texas Southwestern Medical Center Dallas, Dallas, Texas

BACKGROUND: Although the reciprocity hypothesis (that trainees have a tendency to modify evaluations based on the grades they receive from instructors) has been documented in other fields, very little work has examined this phenomenon in the surgical residency environment. The purpose of this study was to investigate the extent to which lenient-grading faculty receive higher evaluations from surgery residents.

METHODS: Evaluation data from 2 consecutive academic years were collected retrospectively at a large university-based General Surgery residency program. Monthly faculty evaluations of residents (15 items) and resident evaluations of faculty (8 items; 1 = never demonstrates, 10 = always demonstrates) were included. Correlation and regression analyses were conducted with SPSS version 22 (IBM; Chicago, IL).

RESULTS: A total of 2274 faculty assessments and 1480 resident assessments were included in this study, representing 2 years of evaluations for 32 core faculty members responsible for completing all resident evaluations and 68 PGY1-5 general surgery residents. Faculty (63% men, 13.5 ± 9.8 years out of training) represented 5 different divisions (general surgery, surgical oncology, transplant, trauma critical care, and vascular) within the general surgery department. Faculty received an average of 71.1 ± 33.9 evaluations from residents over the course of 2 years. The average rating of faculty teaching by residents was 9.5 ± 0.4. Residents received an average of 21.8 ± 0.5 evaluations with average ratings of 4.2 ± 0.4. Correlation analyses indicated a positive relationship between the average rating received from residents and the number of years since faculty completed training (r = 0.44, p = 0.01). Additionally, a significant relationship emerged between ratings received from residents and ratings given to residents (r = 0.40, p = 0.04). Regression analyses indicated that when both variables (years since training, ratings given to residents) were included in the model, only ratings given to residents remained a significant predictor of evaluation ratings received from residents ($F_{(1,32)}$ = 4.40, p = 0.04), with an R^2 of 0.16. Sex or division affiliation did not account for any unique variance.

CONCLUSIONS: These findings suggest that a reciprocity effect exists between surgery faculty and resident evaluations. This effect warrants further exploration, such that efforts to mitigate the risks of providing inaccurate assessments may be developed. Providing trainees with accurate assessments is particularly important given the high-stakes use of these data for milestones, promotion, and graduation purposes, which currently do not account for this reciprocity effect.

SUMMARY: Results suggest that there is a reciprocity effect in the faculty and resident evaluation process. (J Surg Ed ■:■■■-■■■. © 2016 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: evaluation, assessment, residency, reciprocity

COMPETENCIES: Practice-Based Learning and Improvement, Systems-Based Practice

INTRODUCTION

According to the Accreditation Council for Graduate Medical Education (ACGME) program requirements for general surgery, program directors are responsible for ensuring an appropriate educational environment conducive to educating residents in each of the ACGME competency areas.[1] One way in which this is achieved is through regular evaluation of program faculty by trainees. At many institutions, these evaluations of faculty are also regularly used to make pay, promotion, and tenure decisions. Given the potential high-stakes implications of these ratings, the mechanisms underlying this process need to be more closely examined.

Work in other educational domains has noted a "reciprocity effect" in the faculty-student evaluation process, such that a positive association exists between grades that an instructor gives and the evaluations students give to the instructor.[2-4] The premise of the reciprocity hypothesis lies in the idea that trainees have a tendency to modify evaluations based on the grades and feedback that they individually receive; in essence, trainees reward instructors who reward them and punish those who punish them.[5] This relationship has implications for the validity of the evaluation process of surgery faculty and trainees. If this grade-evaluation effect does exist, it suggests that faculty could "buy" good evaluations by supplying lenient evaluations irrespective of actual behaviors and performance. Unfortunately, this phenomenon would also minimize the relationship between performance and trainee assessments, leaving program directors with little more than inflated evaluations of residents from which to make promotion and competency judgments. Although the reciprocity effect has been documented in other fields, very little work has examined this phenomenon in the surgical residency environment.

To further investigate if lenient-grading faculty generally receive higher evaluations, we examined evaluation data from a large general surgery residency program.

METHODS

A retrospective review of 2 years of faculty evaluations of residents and resident evaluations of faculty was conducted. The 2012 to 2014 timeline of this review was chosen to provide the most recent data that coincided with a recent departmental change in evaluation forms and information management systems.

The resident evaluation of faculty assessment tool consists of 8 items reflecting the attending's availability, communication, intraoperative teaching, professionalism, commitment to education, level of clinical knowledge, interest in trainee development, and overall teaching ability rated on a 1 to 10 Likert scale (1 = never demonstrates this quality or trait and 10 = always demonstrates this trait). Residents are similarly evaluated along the 6 ACGME competencies of medical knowledge, patient care, professionalism, practice-based learning and improvement, systems-based practice, and interpersonal and communication skills on a 1 to 5 Likert scale. Evaluations are delivered on a monthly basis to the primary faculty member to rate all residents who rotated on that service. Faculty are strongly encouraged to complete the resident evaluation within 2 weeks of service completion and results are sent directly to the trainee. Residents, on the other hand, rate all primary faculty with whom they worked throughout the month. Based on prior recommendations,[6] they anonymously complete the evaluation at any time and results are provided to each faculty member in aggregate form quarterly. Faculty completed evaluations of residents are not anonymous.

Statistical analyses were conducted with SPSS version 22 (IBM; Chicago, IL). A significance level of $p < 0.05$ was chosen to examine significance of Pearson correlations and linear regression analyses.

RESULTS

A total of 1480 resident assessments and 2274 faculty assessments were included in this study, representing 2 years of evaluations for 68 PGY1-5 general surgery categorical residents and 32 core faculty members responsible for completing evaluations.

Core faculty members consisted of 63% men, were an average of 13.53 ± 9.80 (range: 2-38) years out of training (including fellowship), and represented 5 different divisions (general surgery, surgical oncology, transplant, burns trauma critical care, and vascular) within the general surgery department. Faculty members received an average of 71.1 ± 33.9 evaluations from residents over the course of 2 years. The average rating of faculty teaching by residents was 9.5 ± 0.4. Residents received an average of 21.8 ± 0.5 evaluations with average ratings of 4.23 ± 0.4.

Correlation analyses indicated a relationship between the average rating received from residents and the number of years since faculty completed training ($r = 0.44$, $p = 0.01$), such that faculty who have been in practice for longer periods of time receive higher (better) ratings from residents. Additionally, a significant relationship emerged between ratings received from residents and ratings given to residents ($r = 0.40$, $p = 0.04$), such that faculty who rate residents higher (better) also receive higher ratings from residents. However, the relationship between years since training and average ratings given to trainees was not significant, revealing that the aforementioned ratings given ratings received correlation is not accounted for solely by years in practice.

Regression analyses were conducted to examine how years since training and evaluations of residents influenced ratings received from residents. When both variables (years since training, ratings given to residents) were included in the model, only ratings given to residents remained a significant predictor of evaluation ratings received from residents ($F_{(1,32)} = 4.40$, $p = 0.04$), with an R^2 of 0.16. R^2 reflects the goodness of fit of the model and describes how well the regression line approximates the real data points. This value indicates that 16% of variance between the variables can be accounted for by this model. Sex or division affiliation did not account for any unique variance in these models. The final regression equation is displayed in the Figure.

DISCUSSION

The results of our study confirm the hypothesis that a reciprocity effect exists in the evaluation process for surgery