From feedback to action: Physicians’ teaching performance in residency training
van der Leeuw, R.M.

Citation for published version (APA):
van der Leeuw, R. M. (2013). From feedback to action: Physicians’ teaching performance in residency training

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)

Download date: 27 Dec 2018
Chapter 6

Effects of Residents’ Feedback on Teaching Performance Improvement of Faculty Working in Residency Training: A Longitudinal Study

Renée M. van der Leeuw
Benjamin C.M. Boerebach
Kiki M.J.M.H. Lombarts
Maas Jan Heineman
Onyebuchi A. Arah

Under review
Abstract

Importance
In residency training, teaching faculty are responsible for the preparation of the future workforce while ensuring today’s patient safety. To gain insight into one’s teaching performance and identify areas for improvement, it is essential for faculty to receive feedback from others and engage in guided self-reflection. However, there is a lack of evidence on the effects of these strategies on subsequent teaching performance.

Objective
We investigated the effects of residents’ numerical and narrative feedback and faculty’s completion of a self-evaluation on faculty’s teaching performance improvement.

Design, Setting, and Participants
Prospective multicenter cohort study conducted in The Netherlands between September 2008 and February 2013. Nine hundred and one residents and 1,068 faculty of 65 teaching programs in 16 hospitals were invited to annually (self-) evaluate teaching performance using the validated, specialty-specific System for Evaluation of Teaching Qualities (SETQ).

Main Outcome Measures
Primary outcome was improvement of faculty’s teaching performance as perceived by residents in subsequent evaluations. Multivariable adjusted generalized estimating equations were used to analyze the effects of (i) residents’ numerical and (ii) narrative feedback and (iii) whether or not teaching faculty self-evaluated on faculty’s subsequent teaching performance improvement.

Results
The average response rate over three years was 69% for faculty and 81% for residents. Higher numerical scores were associated with improved teaching performance one year following the first measurement wave (regression coefficient, $b$: 0.077; 95% CI: 0.002 to 0.151; $p=0.045$), but not after the second wave. Receiving more narrative suggestions for improvement was associated with improved teaching performance throughout follow-up. Neither the number of positive comments nor participation in self-evaluation had impact on teaching performance improvement.

Conclusions and Relevance
Evaluation systems for clinical teaching improvement appear helpful in enhancing teaching performance in residency training programs. Previous higher numerical feedback (indicating higher teaching performance) as well as receiving more narrative suggestions for improvement may have led to faculty improvement and being perceived as having improved their teaching performance in subsequent years.
Introduction

In residency training, or (post)graduate medical education, teaching faculty supervise residents to ensure patient safety in training situations. To support faculty in providing high quality training during daily patient care, teaching evaluation systems which provide faculty with performance feedback were developed. Generally, these evaluation systems consist of residents’ feedback, a self-evaluation and a follow-up discussion. Despite the importance of excellent teaching performance of faculty providing residency training, there is limited evidence on the effect of these systems on faculty’s future teaching performance. Organizational psychology literature and knowledge from clinical performance improvement of physicians suggest that research should focus on how individuals improve their performance rather than the question whether performance improves through feedback. Therefore, the aim of the study is to investigate which aspects of teaching evaluation systems predict future performance of individual faculty providing residency training.

First, in teaching evaluation systems residents’ feedback is gathered, which generally consists of numerical scores and narrative comments. The predictive value of numerical scores on subsequent performance was found to be small, but potentially meaningful. However, there is a lack of knowledge on the impact of the direction of the numerical scores (higher or lower) on subsequent performance. Second, in addition to numerical scores, there is increased attention for narrative comments, or written feedback. Narrative comments are regarded as a rich source of information and shown to add individualized information to the evaluation scores. The hypothesis would therefore be that receiving more narrative comments stimulates performance improvement. Third, besides the input of residents’ feedback on teaching performance, a critical self-evaluation of faculty, as is frequently used in multi-source feedback assessments in clinical performance assessment, can complement performance assessment. Comparing both evaluations (residents’ and self) can create direction for improvement: discrepancies between self- and other-ratings can be motivating when over raters receive critical feedback. Furthermore, individuals who increase the discrepancy of self-other ratings by setting higher goals were found to be more motivated to improve their performance. We hypothesize that faculty who complete a self-evaluation and thus create something to compare residents’ feedback to, will subsequently improve their performance more than faculty who do not complete a self-evaluation. It is important to test this hypothesis, because the self-other discrepancy can harm one’s self-image or result in decreased motivation when unfavorable feedback is received by someone with low self-esteem. Therefore, the goal of this prospective cohort study is to investigate the predictive value of numerical scores, the number of both positive comments and suggestions for improvement, and completing a self-evaluation on subsequent teaching performance of faculty.
Methods

Study setting and the SETQ system

In the Netherlands, residency training consists of three to six years training in both university medical centers as well as university affiliated community-based teaching hospitals. Within residency training, all faculty are responsible for the training of residents. The development of the System for Evaluation of Teaching Qualities (SETQ system) started in 2005 with a pilot study followed by implementation of a web-based system to carefully assess all aspects of the system. In 2008, the web-based SETQ system was launched to provide easy access to all faculty interested in evaluating and possibly improving their teaching performance. The system has been validated for anesthesiology, obstetrics and gynecology, surgical specialties, and medical specialties. Residents’ evaluation of faculty’s teaching performance and faculty’s self-evaluation both consist 23 to 30 specialty specific questions. The questions can be organized into five teaching domains: creating a positive and safe learning climate, keeping a professional attitude towards residents, communicating about learning goals, evaluating residents’ competence and performance, and providing feedback. In addition, residents are encouraged to provide narrative comments on faculty’s teaching performance: ‘Please provide strengths of faculty’s teaching performance’ and ‘Please provide concrete suggestions for improvement of faculty’s teaching performance’. Residents are free to choose whom to evaluate, based on whom they have worked with. After completion of the measurement instruments, faculty individually receive an automatically generated SETQ feedback report containing all residents’ evaluations, their self-evaluation scores, a numerical comparison with colleagues, a peer group ranking (in groups with 15 or more faculty), and all narrative comments. Previous studies showed that the feedback reports of faculty who received four to six residents’ evaluations could be considered statistically reliable (Cronbach’s alpha >0.70). After the first wave of measuring teaching performance, receiving feedback and completing follow-up, subsequent measurements were complemented with an additional question on improvement of faculty’s teaching performance: ‘If you worked with this faculty during the past year, have you observed any changes in his or her teaching performance as your supervisor (since the last SETQ evaluation), and if so, how would you qualify these changes?’ This question was answered on a 4-point Likert response scale: 1 = declined; 2 = not changed; 3 = somewhat improved; 4 = greatly improved; and the option “I cannot judge”.

Study population

From September 2008 to March 2011, residents and faculty could enroll in the yearly evaluation of teaching performance of individual faculty through the SETQ system. Participants were followed for three consecutive years (which included two waves of receiving SETQ feedback and two follow-up measurements on teaching performance improvement). In total, 1068 faculty and 901 residents from 65 residency training programs,
representing four university-based and twelve community-based teaching hospitals, were invited. No financial compensation was provided for participation.

**Outcome, Predictors and Covariates**

The main outcome of all the models was the teaching performance improvement of faculty rated by the residents one year after the feedback was provided. Residents’ ratings of improvement on the above mentioned 4-point Likert scale were aggregated to faculty level and used as outcome in the analyses. The predictor variable ‘numerical score’ was calculated for each faculty by taking the average score provided by the residents over all numerical SETQ items. Residents’ written feedback was coded and counted before included in the analyses. We coded the narrative comments on their direction: ‘positive comments’ or ‘suggestions for improvement’ and counted the number of these comments. Counting the number of comments in a specific category was found to be possible with an inter-rater reliability of 0.98. Consequently, one researcher (RvdL) could individually code the direction and number of narrative comments. The number of positive comments and suggestions for improvement were predictor variables in the models. Finally, we also included a binary variable on whether or not each faculty completed a self-evaluation of their own teaching performance concurrently during the wave in which residents evaluated them.

**Data analysis**

Prior to analysis, all variables were examined for accuracy of data entry, missing values, and meeting the assumptions of multivariable analysis. We handled missing data using the well-established techniques of multiple imputation. We created five imputed databases which were then analyzed in parallel; their results were pooled using methods that accounted for the uncertainty of the imputation process. The data were aggregated to the faculty level. We analyzed our data using linear generalized estimating equations (GEE) that allowed for the clustering of the outcome data on faculty within hospitals and specialties. Given our longitudinal data, we specified our regression models as structural nested mean models, the parameters of which were estimated using sequential g-estimation. G-estimation is a sophisticated analytical technique developed for use in longitudinal data settings where some predictors can be mediators and confounders of subsequent predictor-outcome relationships, as in this study. We fit four final models altogether. The first model analyzed the influence of residents’ feedback, namely i) their numerical score, ii) the number of positive comments provided, and iii) the number of suggestions for improvement provided in the first SETQ wave as predictors of perceived performance improvement measured in the second SETQ wave. That is, the outcome was faculty’s teaching performance improvement, as rated by the residents one year later (at wave 2). The second model was equivalent to model 1, but further adjusted for residents’ feedback provided in the second SETQ wave and using faculty’s teaching performance improvement one year after (that is, at wave 3) as the outcome. Models 3 and 4 added, to models 1 and 2 respectively, the variable on
whether faculty performed self-evaluation. All models were additionally adjusted for these covariates: hospital, specialty, faculty’s age, and faculty’s gender. We conducted all analyses using SPSS 20.0 for Windows OS (IBM SPSS Inc., Chicago, 2012).

**Ethical consideration**

We consulted the institutional ethical review board of the Academic Medical Center of the University of Amsterdam (AMC). They confirmed that the Medical Research Involving Human Subjects Act (WMO) did not apply to this study and thus waived ethical approval.

**Results**

In total, residents completed 17,049 evaluations of faculty’s teaching performance. Over the three years, residents’ response rates were 91%, 80% and 72% respectively. After data imputation, data was available for 992 faculty for the first wave and 847 faculty for the second wave. Participants’ characteristics are shown in table 1.

<table>
<thead>
<tr>
<th>Table 1. Participants’ characteristics.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of faculty invited to participate</td>
</tr>
<tr>
<td>Number of residents invited to evaluate faculty’s teaching performance during the first SETQ wave (response rate)</td>
</tr>
<tr>
<td>Number of residents invited to evaluate faculty’s improvement after the first SETQ wave and faculty’s teaching performance during the second SETQ wave (response rate)</td>
</tr>
<tr>
<td>Number of residents invited to evaluate faculty’s improvement of teaching performance after the second SETQ wave (response rate)</td>
</tr>
<tr>
<td>Medical specialties (percentage)</td>
</tr>
<tr>
<td>Surgical specialties (percentage)</td>
</tr>
<tr>
<td>University teaching hospital (percentage)</td>
</tr>
<tr>
<td>University affiliated hospitals (percentage)</td>
</tr>
<tr>
<td>Male (percentage)</td>
</tr>
<tr>
<td>Age of faculty (standard deviation)</td>
</tr>
</tbody>
</table>

As shown in table 2, faculty received an average numerical score of 3.82 over the years. The narrative feedback per year consisted of approximately 10 positive comments and 4 suggestions for improvement. Most faculty completed a self-evaluation (812 in the first year and 597 in the second year). When categorizing residents’ perception of faculty’s improvement, most faculty somewhat improved their teaching performance or did not change.
The results of the multilevel regression models indicate that faculty who received higher numerical scores were more likely to improve ($b=0.077$, $p=0.045$ in model 1 and $b=0.077$, $p=0.042$ in model 3) in SETQ wave 1, however, no effect of numerical scores was found in SETQ wave 2. The number of positive comments was not associated with improvement in both SETQ wave 1 and wave 2. Faculty who received more suggestions for improvement were more likely to improve their teaching performance (first SETQ wave: $b=0.010$, $p=0.056$ for model 1, and $b=0.011$, $p=0.050$ for model 3; second SETQ wave: $b=0.009$, $p=0.006$ for model 2 and $b=0.009$, $p=0.006$ for model 4). Whether or not faculty self-evaluated did not impact their improvement. All results are presented in table 3.

### Discussion

#### Main findings

This study provides empirical data on the effect of residents’ feedback and faculty’s self-evaluation on subsequent teaching performance of faculty working in residency training. Receiving higher numerical scores increased performance, but only after the first feedback wave. The positive effect of the suggestions for improvement on teaching performance was seen in both the first and second feedback waves. Finally, completing a self-evaluation or receiving more positive feedback had no effect on subsequent improvement.

#### Explanation and interpretation of results

Our aim was to investigate the effect of different aspects of teaching performance evaluation systems on teaching performance improvement.
Numerical scores were expected to yield small effect sizes, which was confirmed by this study. Nonetheless, even small effect sizes may represent meaningful performance improvement. Surprisingly, we found that a higher numerical score had a positive effect on improvement in the first year but no effect in the second year. Residents may become more critical of their ‘good’ teachers due to increased attention to teaching performance through participation in the SETQ system. This could have taken away the effect that was found in the first SETQ wave. The suggestions for improvement showed a sustained effect over the years. Faculty’s improved performance in response to suggestions for improvement may be explained by Finkelstein and colleagues’ theory on the response to positive and negative feedback in relation to a person’s level of expertise. They state that novices seek and respond more to positive feedback and experts favor negative feedback as it increases a sense of urgency that their goals are (not yet) attained. In terms of our study findings, it suggests that most faculty are or perceive themselves as experts, as they responded to the suggestions for improvement (‘negative feedback’) in relation to performance improvement. We found no effect of the number of positive comments on faculty’s teaching performance improvement. This may be the result of the above-described effect of experts’ seeking

### Table 3. Sequential g-estimates of the effects of residents’ feedback and faculty’s self-evaluation on faculty’s teaching performance improvement.

<table>
<thead>
<tr>
<th>Wave 1 predictors for wave 2 outcome</th>
<th>Wave 2 predictors for wave 3 outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regression coefficient (b)</strong></td>
<td><strong>95% confidence interval (CI)</strong></td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td><strong>Regression coefficient (b)</strong></td>
</tr>
<tr>
<td><strong>95% CI</strong></td>
<td><strong>p-value</strong></td>
</tr>
</tbody>
</table>

**Model 1**

| Numerical score | 0.077 | 0.002 to 0.151 | 0.045 | -0.016 | -0.074 to 0.042 | 0.595 |
| Number of positive comments | -0.005 | -0.010 to 0 | 0.066 | 0.001 | -0.002 to 0.004 | 0.524 |
| Number of suggestions for improvement | 0.010 | 0 to 0.021 | 0.056 | 0.009 | 0.003 to 0.016 | 0.006 |

**Model 2**

| Numerical score | 0.077 | 0.003 to 0.151 | 0.042 | -0.015 | -0.073 to 0.043 | 0.600 |
| Number of positive comments | -0.005 | -0.010 to 0.001 | 0.085 | 0.001 | -0.002 to 0.004 | 0.530 |
| Number of suggestions for improvement | 0.011 | 0 to 0.021 | 0.050 | 0.009 | 0.003 to 0.016 | 0.006 |
| Self-evaluation (reference = not self-evaluated) | -0.079 | -0.199 to 0.041 | 0.185 | 0.008 | -0.075 to 0.092 | 0.835 |

All models were adjusted for hospital, specialty, faculty’s sex and faculty’s age.
and responding to negative feedback. Although the number of positive comments did not predict performance improvement, the positive comments may be vital in supporting faculty's motivation when receiving negative feedback and preventing a detrimental effect of negative feedback on performance. Completing a self-evaluation did not predict performance improvement. Whether or not faculty complete a self-evaluation might capture something that is not of influence on residents' perception of teaching performance. In other words, the internal reflective process that may have been initiated by completing a self-evaluation may not be within reach of residents' perception and hence not be captured with our improvement question. This does not make residents' perception of improvement an incorrect outcome measure; external assessment is more accurate than self-reported outcomes.

**Strengths and limitations**

The strengths of this study are its longitudinal observation of a study population consisting of a large number of participants across multiple specialties from multiple hospital and medical training sites. Furthermore, the statistical analyses were conducted using sophisticated statistical methods. Multiple imputations were used to deal with missing data in an appropriate way, and sequential g-estimation was implemented to study predictors of changes in participants' teaching performance over time.

A limitation of our study could be the use of residents' perceptions of faculty's improvement as the only outcome measure. Although the perception of residents is preferred over faculty's self-perceived improvement, residents are unable to capture the internal processes that may have been sparked by the feedback. Additional measures such as observation of teaching practice could increase the strength of the main outcome measure. Finally, the exact content of the follow-up actions of faculty is unknown, but this may play a vital role in improving teaching performance. As noted in a clinical multi-source feedback program, self-reported change in practice was found to depend on the perceived quality of a mentor who would encourage reflection.

**Implications for practice**

The results of this study show that even faculty who score very high on their teaching performance can still improve their teaching performance through performance feedback. Even more, the high performers were more likely to improve their teaching performance after the first feedback wave compared to the low performers. Additionally, the suggestions for improvement as phrased by residents are effective aspects of the feedback system. A concrete implication for faculty's teaching practice is thus to pay attention to these suggestions for improvement. Furthermore, residents should provide suggestions for improvement in order to aid faculty's teaching performance improvement. Finally, longitudinal participation appears to be beneficial. As noted earlier in clinical performance feedback, it seems to stimulate participants to take action when they know there will be another evaluation the following year.
Suggestions for future research

Many performance assessment systems include the professionals’ self-evaluation. Based on this study’s findings that self-assessment does not seem to influence performance improvement, exclusion of self-evaluations from performance assessment systems could be suggested. However, further research is needed to investigate the possible internal processes that self-evaluation may stimulate. As noted above, residents’ perception of faculty’s teaching performance improvement might not capture internal processes. Since the understanding of improvement of teaching performance is still in its infancy, it deserves further research before removing self-evaluation from the system. Furthermore, other contextual characteristics should be investigated to increase our understanding of performance and improvement. For example, the way feedback is discussed individually as well as in the group or between faculty and residents. Future work should elaborate on how to measure improvement in other ways than through residents’ perception.
References


