Evaluating clinicians’ teaching performance
Boerebach, B.C.M.

Citation for published version (APA):
CHAPTER 9

GENERAL DISCUSSION

Benjamin C.M. Boerebach
OBJECTIVE OF TEACHING PERFORMANCE EVALUATION

As this thesis is titled *evaluating clinicians’ teaching performance*, we will start this discussion chapter by shortly readdressing for which purpose clinicians’ teaching performance evaluations are conducted. In other words, for which purpose has the System for Evaluation of Teaching Qualities (SETQ), that is nowadays used to evaluate the teaching performance of over 5000 clinician teachers, been developed?

As stated in the introduction of this thesis, the SETQ has been developed as a formative evaluation system.\textsuperscript{1,2} In formative evaluation systems, clinicians use evaluations to generate a critical appraisal of their own performance for their personal development. In the case of the SETQ that is for their development as clinician teachers. Clinicians participate in a formative evaluation system driven by an intrinsic motivation to gather performance data and to improve performance when possible.\textsuperscript{3,4} In a formative context, it is assumed that clinicians are willing to engage in continuous evaluations and try to respond accurately to the performance feedback after each evaluation.\textsuperscript{3-6}

To facilitate clinicians in generating an appraisal of their performance data, structured, valid and reliable external performance data have to be available.\textsuperscript{4} Because external performance data are often not easily available for clinicians,\textsuperscript{4,7} the SETQ aims to generate such data and report it back in a structured report to clinicians. The SETQ tools are designed to yield rich, detailed and specific performance information that can help clinicians in the process of informed self-assessment (model is described in the introduction of this thesis) aiming to further develop their teaching qualities. Besides, the SETQ includes a self-evaluation aimed to facilitate the explication of clinicians’ perception of their own performance, which is also believed to facilitate informed self-assessment.\textsuperscript{4}
OVERALL FINDINGS

We will first shortly summarize the findings regarding the two overall aims of this thesis. The first aim was to better understand how we can validly and reliably evaluate clinicians’ teaching performance and changes in their teaching performance over time. To summarize our findings, we found that the SETQ tools appear to be valid and reliable for gathering clinicians’ teaching performance data from trainees and clinicians’ self-evaluation data. After two evaluation cycles that included gathering performance data, reporting that data to clinicians and an individualized follow-up, clinician teachers enhanced their performance scores. Also, trainees indicated that clinicians improved their overall teaching performance after the first, and even more after a second evaluation cycle. However, the improvement in teaching performance scores and trainees’ perception of overall teaching performance improvement did not correspond on clinician level. Specifically, low correlations between these two measures of performance change were found. Among other arguments, the findings of our research lead us to the conclusion that for robust assessments of performance change, multiple metrics, statistics and data sources should be included.

The second aim was to explore if and how trainees’ perception of clinicians’ teaching performance could impact clinicians’ role modeling. The results of our research suggest that clinicians can substantially enhance their role modeling by improving their teaching performance. Overall, displaying a good professional attitude towards trainees seems to impact role modeling most. The influence of specific teaching domains on clinicians’ role modeling differs across specialties. The associations between specific domains of teaching performance and clinicians’ role modeling also hold under different (plausible) assumptions of causality and confounding, although the strength of the associations differs.

In the remaining part of this chapter, we will elaborate on the interpretation of our findings in the context of the existing literature, discuss the implications of our findings for practice, provide suggestions to improve future practice and policy, and identify topics that require further research. This chapter is organized by content. First we discuss the teaching performance data used
in teaching performance evaluations. Second, we elaborate on how these teaching performance evaluations can lead to teaching performance improvement and third, if and how teaching performance evaluations can be used for non-formative purposes. Fourth, we reflect on the advantages of using a programmatic evaluation approach. We will conclude with discussing our findings regarding the impact of clinicians’ teaching performance on their role modeling.

**CLINICIANS’ TEACHING PERFORMANCE DATA**

Evaluations of clinicians’ teaching performance can involve multiple methods, measures and data sources. In our research we focused on trainee-evaluations and self-evaluations of clinicians’ teaching performance using the online SETQ tools. We will first discuss the characteristics of the performance data yielded by the SETQ tools and subsequently discuss the possibility to include other teaching performance data, using other evaluations methods or other data sources.

*The SETQ evaluations*

In chapter two and three of this thesis, two psychometric studies describe the characteristics of the performance data yielded by both the SETQ trainee-evaluations and the SETQ self-evaluations. Based on generalizability analysis, four to seven trainee-evaluations are needed to obtain a reliable evaluation of clinicians’ teaching performance, depending on the specialty and performance domain evaluated. Additional analyses - presented in chapter 3 - showed that teaching performance evaluations based on two or three SETQ completed trainee-evaluation forms could also be considered statistically reliable. This finding suggests that the SETQ tools are efficient in gathering performance data from trainees, also when only a few trainees are available to provide performance data. This ensures the feasibility to gather reliable performance data through the SETQ, even in small training programs. Studies on other teaching performance evaluation tools found that eight or more trainee-evaluations were needed to obtain reliable performance
data, which make those tools less suitable to evaluate clinicians’ teaching performance in small training programs that include only a few trainees.\textsuperscript{13-16} 

In the teaching competency framework published by Srinivasan et al. (2011), many important competencies for clinician teachers have been described.\textsuperscript{17} Some competencies are important for all clinician teachers, some are more purpose-, setting- or context-specific. In the SETQ tools, we included 20 core items for all specialties, and some optional specialty-specific items that can be included for some specialties. According to a systematic review of other teaching performance evaluations tools, tools include one up to 58 performance items.\textsuperscript{18} To our knowledge, none of these other tools capture all the proposed competency domains, probably because of the necessary trade-off between measurement length (longer tools are less feasible in practice) and completeness. It does not seem feasible to evaluate all competencies described in Srinivasan’s competency framework; the same is true for the SETQ.\textsuperscript{17} However, in close collaboration with trainees and clinicians we aimed to construct an important subset of items to evaluate clinicians’ teaching performance on important teaching domains, keeping the trade-off between measurement length and completeness in mind.\textsuperscript{1;2;11} Compared to other evaluation tools, when it comes to psychometric properties, measurement length and number of evaluations needed for reliable performance scores, the SETQ tools appear to be competitive and efficient in gathering clinicians’ teaching performance data.

The research that is included in this thesis focused on the quantitative performance data generated by the SETQ tools. In addition to the quantitative data, the SETQ tools request trainees to provide narrative comments. Research showed that narrative comments are appreciated in addition to numerical data because they allow for more specific and detailed feedback.\textsuperscript{19;20} Because the numerical and narrative performance data are integrated in the SETQ evaluations, (the effects of) both sorts of performance data should be considered jointly. We will shortly summarize the findings of two studies that explored the characteristics of the narrative SETQ comments.\textsuperscript{21;22} The first study focused on the frequency of narrative comments and found that, on average, clinicians receive thirteen positive comments and five sugges-
tions for improvement in their feedback report. Clinicians with high numerical performance scores receive more positive comments, and vice versa, lower scoring clinicians receive more suggestions for improvement. These results suggest that clinicians receive the type of comments they need most, i.e. positive comments to explain how clinicians can maintain high scores and suggestions for improvement on how to enhance lower scores. The results of the second study on narratives indicate that most comments were made about teaching skills and teaching attitudes, which was, of course, the purpose of the SETQ. The specificity of these comments was overall of good quality. In conclusion, trainees are generous in providing high quality narrative comments accompanying the numerical performance data, which will further benefit the quality of the SETQ performance data for clinicians’ informed self-assessment.

Our research provides evidence supporting the psychometric qualities of the SETQ tools. However, we must underscore the importance of viewing validation as an ongoing process. Measurement tools need to be validated and updated over time for their continued relevant and appropriate use in various contexts and groups. Therefore, the psychometric properties of the SETQ tools need to be critically reviewed, we suggest at least every five years, or more frequently when major changes in postgraduate medical education or medicine occur. Further, evaluation tools always need to be revalidated when they are used in new settings, other countries or for different purposes. As an example, our research group recently updated and slightly revised the SETQ tools to make them suitable for European use. The revised tools, called the SETQ-Smart tools, have been pilot tested in several European countries and the data yielded in those pilots had decent properties, comparable to the properties of the SETQ tools presented in this thesis (study not published yet).

It is also important to expand the validation process of the SETQ by including other validation methods in future studies. Currently, evidence about the convergent, predictive, discriminant and concurrent validity of the SETQ tools is limited. Future research can expand this evidence by relating performance scores to outcomes that they are expected to correlate with. Such outcomes
should reflect teaching and clinical performance, including patient outcomes following care delivered by trainees and attending clinicians. For example, the SETQ scores could be related to outcomes of colleague- or patient completed performance evaluations, guideline adherence data on medical or surgical procedures, and to data on patients’ risk-adjusted quality of life or mortality.

Overall, there is positive validity evidence supporting the use of the SETQ tools to gather clinicians’ teaching performance data, however, there is still an open research agenda to further substantiate the evidence on the SETQ tools.

*Teaching performance evaluations using other data sources*

The SETQ tools and most other teaching performance evaluation tools focus on trainee-evaluations, sometimes combined with self-evaluations, to provide clinicians with data on their teaching performance.\(^{18}\) However, other methods have been used to evaluate clinicians’ teaching performance, including peer evaluations and observations of performance. Because those methods rely on other data sources they will provide performance data from a different perspective and will therefore be complementary to the trainee- and self-evaluations.

Beckman et al. (2003) found that teaching performance evaluations completed by peers yielded reliable data for most teaching domains.\(^ {23}\) An additional study by Beckman et al. (2004) found that the evaluations completed by peers yielded different evaluation scores compared to the same evaluations completed by trainees, i.e. the correlations between the scores given by peers and trainees were low.\(^ {24}\) Although this was a small sized single center study and the only study to our knowledge that compared the data yielded by peer- and trainee-evaluations, the results suggest that teaching performance evaluations based on peer- and trainee-data can yield different evaluation results.\(^ {24}\)

Again another data source can be the direct observation of clinician performance. Such observations by trained observers have been used in two older
studies to evaluate clinicians’ teaching performance, providing evidence that this method is feasible to evaluate specific domains of clinicians’ teaching performance.\textsuperscript{25,26} However, there is little evidence about the feasibility to obtain reliable observational data of clinicians’ overall teaching performance in clinical settings. This absence of this evidence can imply that the feasibility of this method in practice is limited, because successful attempts would probably have been published as they are of interest to many stakeholders in the field of faculty development. Also, to our knowledge, no study compared the results of teaching performance evaluations based on observations with evaluations based on trainee- or peer completed evaluations. In the clinician-patient communication literature, several studies report that the correlations between patient- and observer evaluations of clinicians’ behavior were small.\textsuperscript{27-29} Besides, as has been discussed in chapter 2 and 4 of this thesis, many studies showed that self-evaluations of performance rarely have high correlations with observed performance.\textsuperscript{7,8,30} Although these findings may differ somewhat for clinician-trainee interactions, it shows that evaluations completed by stakeholders who are directly involved in an interaction process (patients, peers and clinicians themselves) do often not match with evaluations based on “neutral” external observations.

In conclusion, the results of evaluations based on a single method of data collection among a homogenous group of feedback providers will provide an incomplete view of clinicians’ teaching performance when used in isolation. Combining several evaluation methods is suggested to make evaluation outcomes more robust.\textsuperscript{31,32} Programmatic evaluations have proven to be effective to evaluate students’ performance in clerkships.\textsuperscript{33,34} It seems very likely that combining several valid and reliable evaluation methods, based on different data sources, will make clinicians’ teaching performance evaluations more robust. Further research is needed to explore the feasibility of such programmatic evaluations of teaching performance. Nowadays, performance evaluations are mostly not directly integrated in a program of evaluation, although other performance data is sometimes included in discussions of the evaluation results.
CHAPTER 9

IMPROVING TEACHING PERFORMANCE THROUGH THE SETQ EVALUATIONS

Many studies explored the effect of performance feedback on subsequent performance. The findings presented in chapter 4 and 5 of this thesis suggest that clinicians can improve their teaching performance after a performance evaluation cycle. This finding is in line with three systematic Cochrane reviews which summarized the effect of clinical audit and feedback (which was defined as: “summary of clinical performance data of health care over a specified period of time”) on clinicians’ subsequent clinical performance. These reviews concluded that audit and feedback can be effective in improving professional practice. The effectiveness of the feedback depends on characteristics of the feedback data, factors within learners, external conditions, sources of information, relationships between people and on factors within the practice environment. In the context of this thesis, factors regarding the feedback data and the evaluation process are of special interest. According to the literature, evaluations are more effective when the content and procedure of evaluations can be adjusted to clinicians’ own learning needs and goals. The feedback data that is yielded by an evaluation should come from sources that are perceived as credible to provide accurate feedback. In addition, the feedback should be clear and specific and should include narrative comments accompanying numerical evaluation scores. Also, the feedback should be presented repeatedly and the evaluation procedure should contain both specific goals and action plans. We expected that the factors impacting the effectiveness of clinical performance evaluations could also impact the effectiveness of teaching performance evaluations. Therefore, the SETQ tools have been designed taking the factors for effective clinical performance evaluations into consideration. Although the SETQ tools have been designed carefully and yield valid and reliable performance data, the use of the tools in practice is equally important for the effectiveness of an evaluation. The findings from the literature, as described above, provide some concrete and specific suggestions for the effective use of the SETQ tools in practice. We will mention and discuss six of them.
First, as with any formative evaluation system, clinician teachers should be engaged in the evaluation process and willing to self-direct their learning in response to the performance feedback.\textsuperscript{3,4} Therefore, it is important to inform clinician teachers accurately and timely about the evaluation procedure, purpose and goals and to engage them during planning, conducting and the follow-up of an evaluation.

Second, the evaluation procedures should be flexible enough to account for learning needs and personal characteristics of clinician teachers such as experience, knowledge and self-confidence.\textsuperscript{39} Currently, some SETQ procedures are flexible and can be adapted to local needs and preferences of training programs. The evaluation period starting date and length, the optional inclusion of specialty specific performance items and methods to discuss the evaluation feedback once the performance data has been gathered can vary between training programs. This allows training programs to combine or attune the SETQ evaluations with other performance evaluations or to adjust them for local goals and needs. The flexibility to adjust the SETQ evaluation procedures for individual clinician teachers within training programs is currently limited. To answer to clinicians’ individualized needs, our research group is designing and experimenting with innovations such as an individualized structured reflection method to facilitate the interpretation of the SETQ performance data and building an online environment that allows for customized feedback reports with specific targets and reference scores for individual clinicians. It is expected that those two developments will enhance the flexibility of the SETQ for individual clinician teachers.

Also, in future evaluations we may experiment with the option for clinicians to focus on specific teaching domains during an evaluation. Future research can explore and experiment with methods to further enhance the flexibility to accommodate individual needs and preferences.

Third, the evaluations should be completed by sources that are perceived as credible by the evaluated clinician teachers. Credible sources are described as people who actually observed clinicians’ performance, understand the specialty domain and have a position of beneficence and non-maleficence.\textsuperscript{5,35,41} Trainees have observed clinicians’ performance, understand the specialty
domain and will probably have a position of beneficence, because the performance feedback they provide can enhance the quality of their own training. In the SETQ, all trainees who worked with a clinician teacher over the last year are eligible to evaluate clinicians’ performance. To further enhance the perceived credibility of the performance data, it can be even more beneficial to let clinician teachers select the trainees they perceive as most credible to complete the evaluations. A study where clinicians could select their own feedback providers showed that self-selection hardly introduced any bias compared to random selection of feedback providers.\(^{42}\)

Fourth, the evaluations should aim to yield specific narrative comments accompanying the performance scores, specifically suggestions for improvement.\(^{19;20;35;40;41}\) The SETQ requests trainees to accompany performance scores with narrative performance descriptions. As described earlier, trainees are generous in providing high quality narrative feedback in the SETQ evaluations.\(^{21;22}\) The number of narrative comments provided through the SETQ was considerably higher than the number provided through a comparable feedback tool in Canada.\(^{43}\) The limited number of numerical feedback items in the SETQ may have stimulated trainees to provide additional narrative feedback. Also, in Canada trainees had to complete far more evaluations compared to the trainees who completed the SETQ, which could have discouraged them from providing additional narrative comments. It is important to monitor whether trainees keep providing narrative comments, especially when the number of clinician teacher evaluations that trainees have to complete or the number of items they have to score increase. The results presented in chapter 5 of this thesis further substantiate the importance of narrative suggestions for improvement, because the number of suggestions received in a feedback report was positively associated with clinicians’ teaching performance improvement.

Fifth, the evaluation should contain both specific goals and action plans. Goal setting prior to an evaluation and planning actions in response to a performance evaluation are activities that are not part of clinicians’ (daily) practice.\(^{20}\) Previous research suggests that a mentor or facilitator can be very helpful in goal setting and action planning.\(^{41;44}\) Overeem et al. found that
the perceived quality of a mentor was the strongest predictor of self-reported change after receiving multisource feedback.\textsuperscript{45} A mentor may also help clinicians to deal with negative emotions resulting from disappointing evaluation results, which may decrease performance improvement, as reported in chapter 4.\textsuperscript{46} Discussing the SETQ evaluation results with a mentor is nowadays included in some training programs, but wider use of a mentor should be stimulated as it can be an important factor for facilitating change.

Sixth, the performance evaluations should be part of a system that repeatedly provides clinicians with performance data. The circular design of the SETQ, in which teaching performance evaluations are usually repeated every year, accommodates the repeated provision of performance data. The finding of chapter 4 that teaching performance scores were only enhanced after the second evaluation cycle (and not after the first) confirms the importance of repeated feedback. Whether the frequency of once a year is the optimal frequency is unknown and deserves further research. It should be noted that a higher frequency will imply a greater time investment of clinicians and trainees and may therefore be less feasible in practice. On the other hand, feedback that is provided timely – i.e. shortly after the observed behavior – is usually perceived as more effective.\textsuperscript{35}

EVALUATING FOR OTHER PURPOSES

Formative evaluations systems, such as the SETQ, have been developed to assist and facilitate clinicians in their own professional development. This thesis shows that the SETQ can be instrumental in facilitating clinicians’ professional development. However, not all performance evaluations have the primary purpose to facilitate professional development. Evaluations can also be used to monitor performance (governance) or to assess if clinicians qualify to obtain or maintain accreditation or certification. These kind of evaluations can be labeled as summative evaluations.\textsuperscript{47} Eventually, both formative and summative evaluations have the goal to maintain or improve the quality of (certain aspects of) healthcare, but the evaluation approaches differ considerably.\textsuperscript{47} Formative evaluations are self-directed and
aim to provide clinicians with rich and various performance data, while summative evaluations are usually externally guided, mandatory and performed to check if clinicians meet certain pre-specified quality standards. In contrast to the rich and comprehensive data aimed for in formative evaluations, the main outcome of a summative evaluation is usually a judgmental rating (insufficient, sufficient, good, excellent) or dichotomous (go/no-go, fit/unfit). Research has shown that formative and summative evaluations require different characteristics to optimize their effectiveness.48-50 Several important factors that contribute to the effectiveness of a performance evaluation, such as its goal, standards, data characteristics and follow-up method, differ substantially between formative and summative evaluations (table 1). Therefore, once an effective performance evaluation system has been developed for formative purposes, it is inherently less suitable for effective summative evaluations, and vice versa.

**Table 1** Differences in evaluation characteristics between formative and summative evaluation systems

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Formative evaluation</th>
<th>Summative evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>Intrinsic</td>
<td>Extrinsic</td>
</tr>
<tr>
<td>Goal</td>
<td>(informed-) Self-assessment of current state to identify professional development opportunities</td>
<td>Check fitness to practice, compare performance with a quality standard</td>
</tr>
<tr>
<td>Frequency</td>
<td>Driven by learning goals and needs</td>
<td>Pre-specified</td>
</tr>
<tr>
<td>Timing</td>
<td>Adjusted to clinicians’ personal scheme</td>
<td>Pre-determined and standardized</td>
</tr>
<tr>
<td>Performance standards</td>
<td>Self-determined and flexible</td>
<td>Pre-specified and standardized</td>
</tr>
<tr>
<td>Consequences</td>
<td>Personalized development plan</td>
<td>Dichotomous outcome (go/no-go, above/below standard)</td>
</tr>
<tr>
<td>Data sources</td>
<td>Sources perceived as most credible in the perception of clinicians</td>
<td>Only sources that can accurately assess if a pre-specified standard is met</td>
</tr>
<tr>
<td>Data characteristics</td>
<td>Detailed, concrete and preferable includes narrative comments</td>
<td>Equally reliable and valid for every clinician, includes numerical data only</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------------------------------------------------</td>
<td>-------------------------------------------------------------------</td>
</tr>
<tr>
<td>Follow-up</td>
<td>Reflection, preferably assisted by a mentor or facilitator to help clinicians in identifying, explicating and planning development opportunities and activities</td>
<td>Additional training or interventions when a desired level is not met; job promotion</td>
</tr>
</tbody>
</table>

This also accounts for the validity evidence of the SETQ that is presented in this thesis. All studies in this thesis were conducted to provide clinicians’ with formative performance data. If trainees would have known that the performance data they provided through the SETQ would be used for summative purposes, they might have provided different, potentially biased or flawed ratings. Consequently, the validity evidence of tools used to gather performance data for formative use, does not apply when the same tools are used in a summative way. Also, the finding that teaching performance evaluations based on trainee feedback generate different evaluation results compared to peer feedback or observations of performance, as discussed earlier in this chapter, implies that summative performance evaluations should always employ a programmatic approach including multiple data sources and data collection methods. Of course, decisions regarding accreditation, fitness to practice or job promotion cannot be made knowing they are based on incomplete performance data. Furthermore, even when different data sources are included, they have to be combined into one “summary measure” to draw summative conclusions. To achieve this, methods to combine the data should be constructed and choices have to be made about which data source to prefer in case of conflicting outcomes.

In chapter 6 of this thesis we showed that combining and summarizing data for the trainee-evaluations is complex and depends on rather arbitrary choices; combining several data sources and methods will be even more complex.
Although the SETQ is not designed for summative use, summative elements are sometimes added to the formative SETQ evaluations in practice. For instance, in the Academic Medical Center in Amsterdam, discussion of the SETQ feedback reports in annual job reviews with the head of a department is mandatory for all clinicians involved in teaching trainees. Although not intended as a summative review, these job reviews are sometimes experiences or used as a summative activity. Also, at some departments, the SETQ scores of clinicians are compared to nationwide mean scores as a sort of performance norm. As shown in chapter 6, the SETQ scores are complex and situation- and context specific, so uninformed comparisons with nationwide means seem meaningless. Even more, when formative performance scores are wrongly interpreted, it can lead to perverse reactions and eventually discourage clinician teachers from further participation in formative evaluations.

Due to recently changed nationwide regulations, gathering performance data from trainees, using systems such as the SETQ, is now mandatory for all clinician teachers. Adding summative elements to the formative SETQ system is called hybrid use of evaluations. In these hybrid evaluations, apart from the mandatory participation or discussion of the feedback results, the evaluation results are only used for clinicians’ own professional development, as with other formative evaluations. Hybrid use of evaluations is seen more often in the Netherlands. Participation in clinical peer review (called visitatie), and soon in multisource feedback evaluations (known as IFMS), is also mandatory for individual clinicians to obtain or maintain their individual registration as a medical specialist, while the results of such evaluations are only used formatively. The effectiveness of hybrid use is questionable and not theoretically founded. Clinicians who are intrinsically motivated and willing to develop their professional performance will participate in performance evaluations anyway, also on a voluntary basis. On the other hand, performance evaluations will probably remain ineffective for clinicians who are neither willing nor motivated to engage in an evaluation and its intended and necessary follow up activities such as the construction of an action plan. According to the literature, motivating and engaging clinicians in performance evaluations and facilitating them in constructing an action plan when the feedback is
received, seems to be much more effective than making participation in a formative evaluation mandatory.

PROGRAMMATIC PROFESSIONAL PERFORMANCE EVALUATIONS

Programmatic evaluations of clinicians’ teaching performance can combine different data collection methods, data sources and can serve several purposes. Taking this line of thought one step further, we can construct programmatic evaluations of clinicians’ professional performance, which also include their clinical, research and managerial performance. Many clinicians perceive teaching in the clinical context and delivering patient care as an integrated and simultaneous process and find it difficult, sometimes even impossible, to distinguish the two in practice. Even more, there will be few professional skills, attitudes and behaviors that solely impact the quality of patient care, research or education. Most generic skills are equally relevant for teaching, clinical, research and managerial performance. For example providing clear verbal and written communication, displaying a professional attitude, time management and providing clear and constructive feedback are relevant for clinicians’ professional performance, irrespective of whether clinicians are teaching, doing research, providing patient care or managing their practice. So why than evaluate them separately?

In a programmatic approach, clinicians’ professional performance evaluations can be integrated, combined or at least attuned, which makes it easier for clinicians to oversee and engage in all evaluations. Also, the logistical and administrative procedures that come with evaluations will be more efficient when evaluations are integrated in a program, for clinicians, evaluators and administrative personnel.

As described earlier, motivation and engagement are key factors for the effectiveness of performance evaluations. Because we know that poorly attuned performance evaluations with overlap in evaluated performance domains or evaluation logistics decrease the motivation and engagement of clinicians, efforts should be taken to avoid such inefficient procedures.
The first attempts to construct programmatic approaches for clinicians’ professional performance evaluations are currently taken. In a recently established committee, named committee Scherpbier 2.0, experts are discussing how to construct programmatic evaluations in an effective and practical way. In the proposed programmatic approach, promoting and evaluating the internal quality management system of training programs and training regions is stressed instead of dominantly relying on mandatory external quality evaluations. Subsequently, an attempt will be made to integrate such teaching quality management systems with clinical quality management systems. In advance of the recommendations of the Scherpbier 2.0 committee, the institutional governing body for postgraduate medical training at the Academic Medical Center in Amsterdam started to construct a programmatic evaluation to oversee the quality of both postgraduate training and patient care in the next five years. In conclusion, it seems that several important stakeholders perceive the need for programmatic professional performance evaluations. In the near future, we will experience how the evaluation programs work in practice and how they may be improved.

TEACHING PERFORMANCE AND ROLE MODELING

Role modeling is uniformly embraced by the medical and medical education literature as an important and powerful teaching and learning strategy. No less than 90% of medical graduates remembers role models who shaped their professional skills and attitudes.\textsuperscript{53,54} There is ample knowledge about clinical, personal and teaching attributes of clinician teachers that may affect trainees’ perception of clinicians’ role modeling. On the contrary, few empirical studies explored the consequences of role modeling or the changeability of clinicians’ role model status.\textsuperscript{55} We aimed to contribute to the existing literature by the studies described in chapter 7 and 8 of this thesis. We explored if and how clinicians’ role model status could be impacted by their teaching performance. We expected that clinicians could enhance their positive role modeling by improving their teaching performance.
Our expectation was substantiated by the relative strong associations found between the teaching performance domains evaluated by the SETQ and clinicians’ role modeling. Because the retrospective study design did not allow for robust causal inference of the study results, we estimated the associations under different assumptions of causality and confounding in chapter 8. The results show that the associations remain - although they differ in strength - under a variety of plausible causal assumptions, which makes a causal claim between clinicians’ teaching performance and their role model status more robust and more plausible in practice. Thus, our studies provide empirical evidence suggesting that clinicians’ teaching performance can influence their role model status, and that enhancing their teaching performance may enhance their role modeling.

Many teaching performance domains are cognitive in nature, thus they can be learned and improved by clinicians. Even more, in chapter 4 and 5 of this thesis, we provide empirical evidence of clinicians teaching performance improvement. This is good news for clinician teachers who aim to enhance their positive role modeling. In the preceding part of this discussion chapter, we discussed that well designed and appropriately used performance evaluations, can assist clinicians in improving their teaching performance. This improvement in teaching performance can then consequently enhance clinicians’ role modeling.

For clinicians who want to enhance their role modeling, there are more possibilities than through improving their teaching performance. According to Mann (2014), faculty development focused at clarifying the meaning of role models and the process of role modeling can also contribute to clinicians’ role modeling. She also highlights the importance of the institutional environment in supporting positive role modeling. Since the term role modeling is included by competency frameworks such as the CanMEDS, the interest of clinicians teachers in the concept of role modeling has increased. For instance, in the Dutch regulations for postgraduate medical education, it is stated that all clinician teachers should act as good role models. Over the last years, we faced multiple clinicians who experienced difficulties with this obligation. They felt they had insufficient knowledge about what positive role
modeling entailed, how it was related to the quality of their teaching, and most important, how clinicians could enhance their role modeling. Following the suggestion of Mann (2014), we organized several (conference) workshops over the last years aiming to inform clinicians about role modeling and experienced positive reactions on behalf of the participating clinicians. Informing clinician teachers about role modeling and expanding the research on the factors that enhance role modeling and on the consequences of role modeling, will remain important to enhance good role modeling in clinical practice. Future research on factors that enhance role modeling should focus on factors that are cognitive or changeable within a reasonable time period.

Such attributes can include clinicians’ communication with patients, respecting and recognizing the contribution of colleagues and coworkers and managerial- and leadership skills. These may also include institutional interventions aimed to improve faculty development opportunities for clinicians.

Although the specific consequences of positive role modeling require further research, we can speculate about the impact that positive role modeling can have on clinical training. As explained in the first chapter of this thesis, clinicians’ always act as role models when they teach, are observed or work with trainees. Clinicians may be conscious or unconscious of their role modeling and it can occur in the formal, informal or hidden curriculum. Trainees can learn from role models in different ways and in different situations. Shortly, they can learn through conscious imitation of role models, through unconscious imitation of role models, and through reflective appraisal of clinician teachers’ performance. Conscious and unconscious imitation can help students initially to cope with clinical situations or challenges that are new to them. In that case, positive role modeling (i.e. providing good examples and explaining good practice) can directly affect trainees’ clinical practice and will enhance the quality of trainees’ patient care. It has also been suggested that unconscious imitation of behaviors occurs through the hidden curriculum, as part of the socialization of medical trainees at a new department. This is mostly referred to when trainees incorporate inappropriate or unprofessional behaviors recently after they start working at a new department. The enhancement of positive role modeling (and avoiding negative
role modeling) will probably result in trainees incorporating more positive and less inappropriate behaviors, which will eventually benefit the clinical and learning environment and consequently the quality of patient care. Lastly, when trainees learn through reflective appraisal of selected behaviors of clinician teachers, more positive role models will provide trainees with more and probably a greater variety of learning opportunities. This allows them to either learn more or to be more selective in which role models to learn from. Besides, a greater variety of positive role models will allow trainees to select the role models they deem most sufficient and fit for their personal professional development.

In conclusion, positive role modeling can benefit the quality of training in multiple ways. Because the quality of training can impact the quality of patient care, trainees and patients deserve good role modeling by clinicians. To take role modeling seriously can therefore be considered a moral obligation of all clinician teachers.

**CONCLUDING REMARKS**

In this discussion chapter, we explained the overall findings of our research and provided suggestions to incorporate the findings of our research into practice, aiming for more effective performance evaluations in the near future. As is almost inherent to scientific research, every research question answered raises multiple new questions for advancing knowledge. The same accounts for this thesis that identified many opportunities for further research, of which several were phrased in this and previous chapters.

This thesis contributes to the knowledge of teaching performance evaluations and role modeling. Also, it contributes to evidence based (graduate medical) education practice by means of the well-studied quality of the SETQ that is nowadays used by over 5000 clinicians and 5000 trainees. Till now, the number of participants using the SETQ is growing every year. Considering the positive effect that evaluation systems like the SETQ can have on the quality of teaching, and hence, on the quality of patient care, we strongly
feel it is a collective responsibility of the medical profession, researchers, educationalists and policy makers, to continue and further expand this (type of) research. Assisting clinician teachers and trainees in improving their performance serves nothing less than the ultimate goal to deliver high quality care to all patients.
REFERENCES


(9) Boerebach BC, Lombarts K, Arah O. Confirmatory Factor Analysis of the System for Evaluation of Teaching Qualities (SETQ) in Graduate Medical Training. *Eval Health Prof* 2014. in print


(21) van der Leeuw R. *From Feedback to Action*. Amsterdam: University of Amsterdam; 2013.


(49) Elzubeir M, Rizk D. Evaluating the quality of teaching in medical education: are we using the evidence for both formative and summative purposes? *Med Teach* 2002; 24(3):313-319.


(58) Mann KV. *Faculty Development to Promote Role Modeling and Reflective Practice*, in: *Faculty Development in the Health Professions: A Focus on Research and Practice*. 2014.


