Clarifying observation and assessment feedback in workplace-based learning

Els Pelgrim

Clarifying observation and assessment feedback in workplace-based learning

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op dinsdag 24 september 2013
om 12.30u precies

door

Elisabeth Antonia Maria Pelgrim
Geboren op 7 november 1985
te Eindhoven

Promotor

Dhr. Prof. dr. C.P.M. van der Vleuten

Copromotoren

Mw. dr. A.W.M. Kramer

Dhr. dr. H.G.A. Mokkink

Manuscriptcommissie

Mw. Prof. dr. J. de Graaf *(voorzitter)*

Dhr. Prof. dr. W.J.J. Assendelft

Mw. Prof. dr. A.D.C. Jaarsma *(Amsterdam Medical Centre –*
University of Amsterdam [AMC-UvA])

Dhr. dr. J.C.M. Lavrijsen

Mw. dr. M.J.B. Govaerts *(Maastricht University)*

Paranimfen

Mw. J. van der Sanden – Melis MSc

Mw. J.M.R. Huisman BSc

Table of contents

Chapter 1

Introduction

Introduction

The central theme of this thesis is: challenges and opportunities of *workplace-based learning and assessment*. Despite ample research and significant progress in assessment of professional competence[1], many theoretical and practical questions still remain unanswered. The concrete practical question that inspired the studies in this thesis arose in postgraduate training for general practice (GP) in the Netherlands, where an educational evaluation among trainees revealed that trainees were rarely if at all observed during patient encounters[2]. These findings are consistent with reports in the literature about infrequent observation of performance in undergraduate and postgraduate workplace-based training and assessment[3,4] and how this had a detrimental effect on feedback. Learning and assessment in the workplace rely on information from people who, through observation, know about learner behaviour in practice[5]. Indeed, Sargeant showed that feedback recipients questioned the credibility of feedback that was not based on actual observation of their performance[6]. In short, observation is a crucial but too infrequently practised element of workplace-based assessment.

It is also known, however, that feedback, even if based on observation, does not automatically promote learning and performance improvement[4]. Assessment helps learners identify and address their learning needs by providing insight into their performance and generating new knowledge to improve overall performance[7,8], and feedback plays a key role in this. The important questions about feedback are: do feedback providers (trainers or supervisors) provide useful feedback? Do trainees accept feedback and use it to guide their learning and performance improvement?

The above considerations led to the central objective of this thesis: to provide insight into observation of single patient encounters and feedback in workplace-based assessment. This introductory chapter looks at the research on workplace-based learning and assessment. After a broad picture is sketched, the focus narrows to the more detailed research questions. This introductory chapter leads up to the problem statements that were addressed in the studies presented in the following chapters.

Miller's pyramid

To bring clarity to the issue of observation of patient encounters and feedback in assessment of workplace-based learning we start by exploring the assessment literature and investigate characteristics of workplace-based assessment.

In 1990, George Miller presented a classification of clinical performance and its assessment, using a pyramid to illustrate different levels and how they are best assessed (figure 1)[9]. The base of the pyramid is concerned with knowledge: does the learner *know* what is required for professional performance. Knowledge can be

assessed reliably and validly using open and multiple choice questions as well as oral assessments[9,10]. It is also important, however, to *know how* to apply knowledge. Assessment at this level is similar to assessment of knowledge but it is more closely connected with the professional context, for example by the use of patient scenarios and vignettes[1,10]. Next, the *shows how* level moves beyond the merely cognitive level to actual behaviour. What do learners actually do when they have to apply their knowledge? This can be assessed in simulated and standardized situations, such as Objective Structured Clinical Examinations (OSCE). Finally, the top of the pyramid focuses on what a professional *does* when functioning independently in the clinical workplace[9,10]. The topic of this thesis, workplace-based assessment, is situated at the very top of the pyramid.
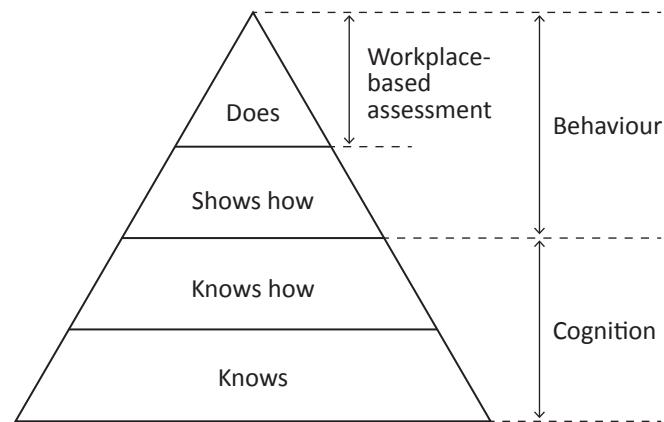

Figure 1 Miller's pyramid showing four layers of assessment

Assessment goals

In addition to Miller's perspective of increasing complexity and authenticity, assessment can be viewed from two goal-oriented perspectives. One goal of assessment is to determine whether a learner has achieved a certain goal, is ready to move to the next level or qualifies for licensure or certification[11]. Termed *summative assessment,* this type of assessment requires systematic recording of learners' achievements and performance[12] to ensure the validity and reliability of high-stakes decisions[12,13]. The second perspective focuses on assessment aimed at enhancement of learning[4,14]. This so-called *formative assessment* is aimed at diagnosing what a learner is doing well and where more work is needed to optimize performance. It affords insight into learners' positive and negative achievements to inform appropriate next steps. Although formative assessment should be valid and reliable as well, it depends strongly on the

quality of the process of seeking, giving, receiving, interpreting, and using feedback. The 'feedback giver' and the 'feedback recipient' are leading actors in this process[14].

How do these two perspectives play out in educational practice? Are they interwoven or are they separate? It has been suggested that summative and formative assessment should be viewed as opposite ends of an assessment continuum[5]. Learners often set learning goals in accordance with summative benchmarks: 'assessment drives learning'[7]. A summative decision can be based on aggregated performance judgements collected over a period of time, i.e. the decision is informed by the sum of formative assessment moments[1,5,12]. In pursuing the objective of this thesis, it is necessary to investigate how the theoretical perspectives of formative and summative assessment can be used to optimize learning and assessment. Should we separate them or integrate them to optimize the benefits of both?

A model of programmatic assessment

A recently developed model of programmatic assessment presents a coherent analysis of the workings of assessment – both formative and summative - in practice[5]. The dual purpose of this model is to maximally support 'on-going learning' as well as robust decisions about progress. The model covers all levels of Miller's pyramid and distinguishes training activities, assessment activities and supporting activities (Figure 2). Training activities (circles) comprise any activity that is conducive to learning, such as tasks (a lecture, a patient consultation, etc.) or artefacts (a project report).
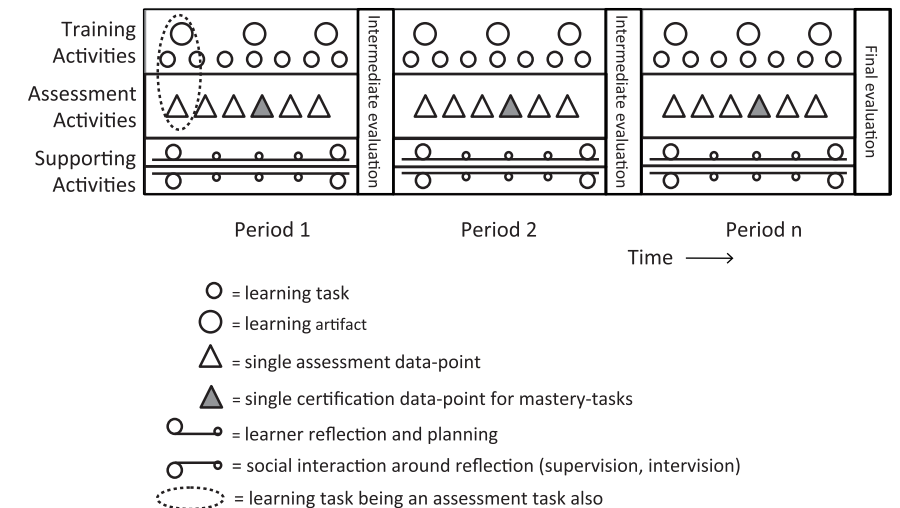

Figure 2 Model of programmatic assessment

Represented as triangles, assessment activities are arranged to afford optimal support of on-going learning. This implies that assessment relies on feedback, subjective information and expert judgement. Also, pass/fail decisions cannot be based on a single data point, except where mastery oriented tasks (black triangles) are concerned, such as resuscitation, which require demonstration of mastery of a specific skill. The majority of assessment tasks, however, are not mastery oriented but developmental assuming a gradual progression of competence (blank triangle).

There are two types of supporting activities: reflection on information from learning and assessment activities and social interaction to support reflection. These activities reflect the notion that feedback has to be interpreted before it can be used to plan new learning tasks or goals and that supervision and 'intervision' (peer supervision) activities should scaffold self-directed learning by promoting and encouraging learners to engage in reflection and planning.

At the end of a training period (Figure 2), all the information that has been collected is assessed in an intermediate progress evaluation. This requires expert judgement to aggregate information across data points. After an appropriate number of training periods, a final evaluation provides information to inform a high-stake decision about learner progress with major consequences for the learner. This decision too relies heavily on expert judgement and requires even more stringent procedural safeguards than intermediate evaluations.

Synopsis

In the preceding we have sketched a broad theoretical framework for workplace-based learning and assessment showing that this type of assessment represents the 'does' level of Miller's pyramid, that it can be both formative and summative or both, and that, from the perspective of programmatic assessment, training, assessment and supportive activities should be combined to ensure optimal evaluations. The model of programmatic assessment does not separate the different goals of assessment (formative versus summative) but attempts to combine them.

How to put these ideas into practice? How to relate the broad objective of this thesis to the practical question that first triggered the research? Where in the model of programmatic assessment are observation and feedback positioned? An analysis of the model reveals that observation can be viewed as a training activity and as an assessment activity, while a feedback discussion after an observation can be seen as a supportive activity. Observation thus seems to be one of the key components of learning and assessment. This brings us back to observational assessment in work-place-based learning: 1) medical students and postgraduate trainees report that observation of performance is rare; 2) feedback does not automatically lead to learning; 3) assessment in workplace-based learning depends on the quality of the

knowledge of feedback providers and on their ability to give useful feedback. We will now take a further look at the role of observation and feedback to bring the objective of this thesis sharply into focus.

Observation

Information for single assessment data points and for feedback is best obtained from direct observation of performance. At all levels of medical education and training observation usually involves learner performance during patient encounters. Within programmatic assessment (Figure 2), observation followed by a feedback discussion can be seen as a training activity, an assessment activity, and a supporting activity. From a summative perspective, it can validate a trainer's impression of a learner's abilities[1,3,4,15] and from a formative perspective it can be the topic of a conversation about the learner's strengths and weaknesses[5]. Obviously, observation is an indispensable component of the workplace as a training setting.

At the centre of this thesis is observation of learner performance in single concrete clinical situations, such as a patient consultation. This is a core element of doctors' professional practice and consequently, in terms of programmatic assessment, it is one of the essential data points of assessment. Trainers can observe learner performance live or via video registration[1]. With direct observation, learner performance can be discussed and feedback given immediately afterwards[7]. This process can be structured and supported by an assessment tool[16-18]. A typical benefit of this type of observation is that the trainer obtains an accurate view and experiences the atmosphere of the consultation. A typical benefit of video observation on the other hand is the opportunity for learners to discuss the consultation with their trainer while simultaneously observing their own performance[9]. Also, trainers can select fragments they wish to accentuate or play the video without interruption to focus on a particular chain of events[19-24]. For learners, both types of observation enable the identification of learning goals. Ultimately, observation of performance can contribute to better patient care[25].

Feedback

Feedback should preferably be a component of any assessment, whether formative or summative. Although research suggests that feedback can be a powerful tool to change learners' behaviour[4,26-28], it should be noted that several studies have shown that learners do not benefit from feedback that is restricted to marks[1,26,29,30], and that feedback should preferably be narrative and detailed, identifying areas where more work is needed. Narrative feedback has been reported to result in higher satisfaction[31] and potential improvement of in-training evaluation[32], and specific feedback appears to be more readily assimilated and have a significant positive influence on performance[30,33,34]. As stated earlier, feedback is characterized by judgements made by people who are knowledgeable about the recipient's performance

and feedback can be supported by checklists, rating scales, rubrics and narrative comments[1].

Important elements, key actors and issues in observation

Trainer and learner are key actors in observations of performance. Like feedback, observation can be supported by assessment tools. In the following we investigate important elements and key actors in observed performance: the *assessment tools*, the *trainer* and the *trainee*. The widespread use of assessment tools intended to encourage observation of trainee performance and support observation of patient encounters in workplace-based assessment[1,4,17], begs the question of the paradox of the reported paucity of observation of trainee performance despite the presence and usage of suitable tools?[2-4] Other questions are how such instruments can provide data points for programmatic assessment and whether they result in enough specific feedback?

Given the importance of their judgements and feedback, trainers are clearly key actors in observation-based assessment. By studying trainers' views and behaviours we may be able to understand and overcome deficiencies. The literature suggests that it is important to optimize subjective information and judgements from trainers before relying on them for assessment[5]. Trainers should be experts in assessment. According to Crossley and Jolly, we need to understand and appropriately utilize the cognitive schemas of trainers[35]. What underpins these schemas? What about the paucity of observation and the shortcomings in the use of feedback for learning and improvement? Can we optimize trainer competence to overcome these deficiencies? The other key actor in workplace-based assessment is the trainee. Teunissen et al. showed that trainees are not passive recipients of feedback, but show active feed-back-seeking behaviour depending on individual and situational variables[36]. Trainee-related factors may affect the use of observation and studying trainees may provide insights that can help to explain and overcome deficiencies. In the complex setting of workplace-based assessment trainees are required to seek, receive, interpret and use feedback. Reflection is commonly advocated as a powerful tool for trainees to enhance the use of feedback, and Sargeant et al. have shown that reflection can help overcome negative feelings that potentially interfere with the acceptance of feedback[37-38]. Reflection and reflective practice are described as essential attributes of a competent health care professional[39]. But do trainees reflect on their performance to optimize the assessment process? And do these reflections have added value?

This combination of elements and actors has resulted in the main questions that are addressed in this thesis:

- What is the role of an assessment tool in feedback processes based on observation of single patient encounters?

- What is the role of the trainee in feedback processes based on observation of single patient encounters?
- What is the role of the trainer in feedback processes based on observation of single patient encounters?

Sub-questions following from the main questions are discussed in the following chapters. In the general discussion the findings are synthesized in light of the three main questions.

Thesis outline

The central aim of this thesis is to clarify issues related to observation and feedback in workplace-based assessment. The actors in the complex setting of workplace-based learning and assessment are the *trainee,* who needs to develop clinical competence, the *trainer,* who guides and assesses this development, and the setting of real patient care which cannot (and should not) be fully structured or standardized for assessment purposes, but which should be modified to ensure fair assessment and useful feedback, possibly by the use of an *assessment tool*. To unravel the process of work-place-based learning and assessment we first conducted a literature review of assessment instruments in use for observation of patient consultations. From the perspective of optimizing the assessment process, we examined the characteristics and key qualities of assessment instruments, such as reliability, validity and educational effects (Chapter 2).

The research described in the following chapters was conducted in postgraduate training in general practice (GP) in the Netherlands, where trainees work in a general practice supervised by a GP trainer. This one on one setting allowed us to examine the relationship between trainer and trainee in relation to observation of and feedback on trainee performance and the learning of trainees.

Chapter 3 describes a study in which trainees were interviewed about the use of observation and feedback in practice, the role of the trainer and the assessment tool in this, and the use of assessment to promote learning.

In chapter 4 we analyzed completed forms of an assessment instrument to determine the quality of reflection and feedback and to investigate the interplay between assessment form, trainer and trainee.

Chapter 5 examines whether it is the reflections of the trainee and/or the feedback from the trainer that affects the learning of trainees. The interplay of assessment form, trainer and trainee are examined.

Chapter 6 focuses on the trainer as a key actor, examining whether or not attributes of the trainer (personality, self-efficacy and interpretation of training tasks) influence the way feedback is given in observational assessment.

Chapter 7 presents the general discussion of the research in light of the three main questions. We weigh the strengths and limitations of the research and consider directions for future research as well as the implications of the findings for work-place-based learning and assessment.

Because this thesis is based on published journal articles, some repetition across chapters is inevitable.

References

1.   Vleuten, van der, C.P., Schuwirth, L.W., Scheele, F., Driessen, E.W. & Hodges, B. (2010). The assessment of professional competence: building blocks for theory development. *Best Practice & Research Clinical Obstetrics & Gynaecology, 24,* 703 – 719.
2.   Velden, van der, L.F.J. & Batenburg, R.S. (2011) De opleiding tot huisarts opnieuw beoordeeld. Een onderzoek onder huisartsen in opleiding en alumni. http://www.nivel.nl/pdf/rapport-kwaliteit-huis-artsenopleiding.pdf. [Retrieved: February 2013].
3.   Fromme, H.B., Karani, R. & Downing, S.M. (2009). Direct observation in medical education: a review of the literature and evidence for validity. *Mount Sinai Journal of Medicine, 76,* 365 – 371.
4.   Norcini, J.B. & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher, 29,* 17.
5.   Vleuten, van der, C.P.M., Schuwirth, L.W.T., Driessen, E.W., Dijkstra, J., Tigelaar, D., Baartman, L.K.J. & Tartwijk, van, J. (2012). *Medical Teacher, 34,* 205 – 214.
6.   Sargeant, J., Mann, K. & Ferrier, S. (2005). Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Medical Education, 39,* 497 – 504.
7.   Epstein, R.M. (2007). Medical education: Assessment in medical education. *The New England Journal of Medicine, 356,* 387 – 396.
8.   Schuwirth, L. & Vleuten, van der, C. (2004). Merging views on assessment. *Medical Education, 38,* 1208 – 1210.
9.   Miller, G.E. (1990). The assessment of clinical skills competence performance. *Academic Medicine, 65,* S63 – S67.
10. Wass, V., Vleuten, van der, C., Shatzer, J. & Jones, R. (2001). Assessment of clinical competence. *Lancet, 357,* 945 – 949.
11. Dauphinee, W.L. (2002). *Licensure and certification.* In: Norman, G.R., Vleuten, van der, C.P.M. & Newble, D.I. International Handbook of Research in Medical Education. Dordrecht: Kluwer Academic Publishers, 835 – 882.
12. Harlen, W.J & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education, 4,* 365 – 379.
13. McGill, D.A., Vleuten, van der, C.P. & Clarke, M.J. (2011). Supervision assessment of clinical and professional competence of medical trainees: a reliability study using workplace data and a focused analytical literature review. *Advances in Health Sciences Education, 16,* 405 – 425.
14. Watling, C.J. & Lingard, L. (2012). Toward meaningful evaluation of medical trainees: the influence of participants' perceptions of the process. *Advances in Health Sciences Education, 17,* 183 – 194.
15. Starmer, A.J., Randolph, G.D., Wysocki, K.L. & Steiner, M.J. (2009). Direct observation of resident-patient encounters in continuity clinic: a controlled study of parent satisfaction and resident perceptions. *Education for Health, 22,* 325.
16. Norcini, J.J., Blank, L.L., Duffy, F.D. & Fortna, G.S. (2003). The mini-CEX: a method for assessing clinical skills. *Annals of Internal Medicine, 138,* 476 – 481.
17. Norcini, J.J., Blank, L.L., Arnold, G.K. & Kimball, H.R. (1995). The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Annals of Internal Medicine, 123,* 795 – 799.
18. Donato, A.A., Pangaro, L., Smith, C., Rencic, J., Diaz, Y., Mensinger, J. & Holmboe, E. (2008). *Medical Education, 42,* 1234 – 1241.
19. Hämäläinen, P. (2004). Interactive video Mirrors for sports training. *NordiCHI '04,* 199-202. DOI: 10. 1145/1029014.108044.
20. Star, J.R.S. & Strickland, S.K. (2008). Learning to observe: using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education, 11,* 107 – 125.
21. Ram, P., Grol, R., Rethans, J.J., Schouten, B., Vleuten, van der, C. & Kester, A. (1999). Assessment of general practitioners by video observation of communicative and medical performance in daily practice: issues of validity, reliability and feasibility. *Medical Education, 33,* 447- 454.
22. Ives, J.C.S., Straub, W.F. & Shelley, G.A. (2002). Enhancing athletic performance using digital video in consulting. *Journal of Applied Sport Psychology, 14,* 237 – 245.

23. Marita, P., Leena, L. & Tarja, K. (1999). Nurses' self-reflection via videotaping to improve communication skills in health counseling. *Patient Education and Counseling, 36,* 3-11.

24. McCurdy, B.L. & Shapiro, E.S. (1988). Self-observation and the reduction of inappropriate classroom behaviour. *Journal of School Psychology, 26,* 371 – 378.

25. Holmboe, E.S. (2004). Faculty and the observation of trainees' clinical skills: problems and opportunities. *Academic Medicine, 79,* 16 – 22.

26. Black, P. & William, D. (1998). Assessment and classroom learning. *Assessment in Education, 5,* 68.

27. Black, P. & William, D. (1998). Inside the black box – Raising standards through classroom assessment. *Phi Delta Kappa, 80,* 139 – 149.

28. Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45,* 25.

29. Rolfe, I. & McPherson, J. (1995). Formative assessment – How am I doing? *Lancet, 345,* 837 – 839.

30. Shute, V.J. (2008). Focus on formative feedback. *Review of Educational Research, 78,* 153 – 189.

31. Overeem, K., Lombarts, M.J., Arah, O.A., Klazinga, N.S., Grol, R.P. & Wollersheim, H.C. (2010). Three methods of multi-source feedback compared: a plea for narrative comments and co-workers' perspectives. *Medical Teacher, 32,* 141 – 147.

32. Govaerts, M.J., Vleuten, van der, C.P., Schuwirth, L.W. & Muijtjens, A.M. (2005). The use of observational diaries in in-training evaluation: student perceptions. *Advances in Health Sciences Education, 10,* 171 – 188.

33. Sargeant, J., Mann, K., Vleuten, van der, C. & Metsemakers, J. (2008). 'Directed' self-assessment: practice and feedback within a social context. *Journal of Continuing Education in the Health Profession, 28,* 47 – 54.

34. Archer, J.C. (2010). State of the science in health professional education: effective feedback. *Medical Education, 44,* 101 – 108.

35. Crossley J. & Jolly, B. (2012). Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. *Medical Education, 46,* 28 -37.

36. Teunissen, P.W., Stapel, D.A., Vleuten, van der, C., Scherpbier, A., Boor, K. & Scheele, F. (2009). Who wants feedback? An investigation of the variables influencing residents' feedback-seeking behavior in relation to night shifts. *Academic Medicine, 84,* 910 – 917.

37. Sargeant, J., Mann, K., Sinclair, D., Vleuten, van der, C. & Metsemakers, J. (2008). Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Advances in Health Sciences Education, 13,* 275 – 288.

38. Sargeant, J.M., Mann, K.V., Vleuten, C.P. & Metsemakers, J.F. (2009). Reflection: a link between receiving and using assessment feedback. *Advances in Health Sciences Education, 14,* 399 – 410.

39. Mann, K., Gordon, J. & MacLead, A. (2009). Reflection and reflective practice in health professions education: a systematic review. *Advances in Health Sciences Education, 14,* 595 – 621.

Chapter 2

In-training assessment, using direct observation of single patient encounters, a literature review

In-training assessment, using direct observation of single patient encounters, a literature review

E.A.M. Pelgrim, A.W.M. Kramer, H.G.A. Mokkink, L. van der Elsen, R. P.T.M. Grol, C.P.M. van der Vleuten

Abstract

We reviewed the literature on instruments for work-based assessment in single clinical encounters, such as the mini-clinical evaluation exercise (mini-CEX), and examined differences between these instruments in characteristics and feasibility, reliability, validity and educational effect. A PubMed search of the literature published before 8 January 2009 yielded 39 articles dealing with 18 different assessment instruments. One researcher extracted data on the characteristics of the instruments and two researchers extracted data on feasibility, reliability, validity and educational effect. Instruments are predominantly formative. Feasibility is generally deemed good and assessor training occurs sparsely but is considered crucial for successful implementation. Acceptable reliability can be achieved with 10 encounters. The validity of many instruments is not investigated, but the validity of the mini-CEX and the 'clinical evaluation exercise' is supported by strong and significant correlations with other valid assessment instruments. The evidence from the few studies on educational effects is not very convincing. The reports on clinical assessment instruments for single work-based encounters are generally positive, but supporting evidence is sparse. Feasibility of instruments seems to be good and reliability requires a minimum of 10 encounters, but no clear conclusions emerge on other aspects. Studies on assessor and learner training and studies examining effects beyond 'happiness data' are badly needed.

Introduction

The mini-clinical evaluation exercise (mini-CEX) is widely used for assessment in single work-based encounters of clinical competence at the top of Miller's pyramid – the 'does' level. Currently, assessment of clinical competence is receiving increasing attention, particularly in postgraduate training[1], and assessment of authentic performance is considered the main challenge. Reliable and valid performance measurements that can serve as a gold standard for clinical assessment have as yet not been acheived[1]. Developed for the evaluation of a multitude of clinical competencies[2], the mini-CEX is a single-encounter instrument to be used by professionals in conducting work-based assessment of actual clinical performance. It was originally developed in 1995 in the USA for the evaluation of internal medicine residents' clinical skills[2,3] and its principal characteristics are direct observation of real patient encounters, easy an instant use in day-to-day practice, applicability in a broad range of settings and immediate feedback to the learner after the encounter. These characteristics make the mini-CEX an educational tool that can help learners to gain insight into the strengths and weaknesses of their clinical performance. It can be used to assess multiple competencies, such as communication and professionalism. Typically, the mini-CEX and similar instruments use global assessment scales, provide space for narrative comments and allow for feedback presented by a moderator in a post-encounter review session.

Since the mini-CEX was first introduced, several comparable instruments have been developed for use in undergraduate and postgraduate medical education, including, among many others, 'longitudinal evaluation of performance'[4], 'structured clinical observation[5] and the 'clinical encounter card'[6]. To our knowledge, however, no review has compared the characteristics and key qualities of these instruments. Feasibility, reliability, validity and educational effects are the core elements in determining the utility of assessment methods[7]. The only review of the validity of instruments for work-based clinical assessment was published in October 2009[8]. The authors conclude that many tools are available, but evidence on their validity and descriptions of educational outcomes are scarce. We reviewed the literature on instruments for single-encounter work-based clinical assessment, like the mini-CEX. These instruments appear to hold promise for clinical assessment but too little is known about their characteristics and feasibility, reliability, validity and educational effects.

We addressed the following research questions:
1. What are the similarities and differences between the characteristics of clinical assessment instruments, such as the mini-CEX?
2. What is known about the feasibility, validity, reliability and educational effects of these clinical assessment instruments?

Methods

We conducted two searches of the PubMed database for papers on clinical assessment instruments published before 8 January 2009. For our first search, aimed at identifying papers dealing with the principal characteristics of work-based assessment instruments, we used the following search terms:

- clinical competence (medical subject heading [MeSH] term and text word) OR educational measurement (MeSH term and text word) OR educational measurements (text word) OR clinical skills (text word)

AND

- medical students (MeSH term and text word) OR clinical clerkship (MeSH term and text word) OR internship and residency (MeSH terms) OR internship (text word) OR residency (text word) OR medical education (MeSH term and text word) OR preceptorship (MeSH term and text word)

AND

- observation (text word) OR observe (text word) OR observed (text word).

AND

- feedback (MeSH term and text word)

OR

- reproducibility of results OR feasibility studies OR observer variation OR pilot projects OR psychometrics OR qualitative research OR statistical data interpretation OR Delphi-technique (MeSH terms and text words) OR evaluation studies (MeSH term, publication type and text word) OR validation studies (publication type and text word).

Our second PubMed search was limited to articles published between November 1995 (publication date of the first paper on the mini-CEX) and 8 January 2009, and used the text words:

- mini clinical evaluation exercise OR mini-CEX OR mCEX OR clinical evaluation exercise.

In addition, we manually searched the reference lists of the included articles for relevant articles.

We used the following inclusion criteria:

- the instrument is used by professionals to assess directly observed performance
- the instrument is used in authentic patient encounters
- the instrument uses a generic and global assessment scale
- the instrument allows for feedback immediately after the assessment
- the instrument is used in a postgraduate or undergraduate medical programme

And we applied the following exclusion criteria:

- the instrument is used for peer-, patient-, or self-assessment
- the instrument only assesses technical skills
- the instrument is used in simulated encounters (as opposed to authentic encounters)
- the instrument (only) assesses a 'long case'[9]
- the instrument reports results as a letter or comment
- no abstract available

Articles were selected by four researchers (LvdE, EP, AK and HM). In an initial selection round, two researchers independently selected articles based on the title only. Any disagreements were resolved by discussion. Next, the abstracts of the articles selected in the first round were independently judged by two researchers. Any disagreements on inclusion or exclusion were resolved in a meeting of three researchers. In the final selection round the full text of the remaining articles was read by LvdE or EP.

Data extraction

Data relating to the following characteristics of the assessment instruments were extracted from each article by one researcher (LvdE or EP).

- setting, summative or formative assessment
- type of encounters (e.g. in-patient, out-patient), assessor and learner
- subject of assessment
- rating scale, criteria for the allocation of marks, frame of reference
- the assessment form
- type of feedback (quantitative/qualitative)
- assessor training
- learner instruction

Next, two of four researchers (LvdE, EP, AK and HM) extracted data related to the aspects addressed by the second research question:

- feasibility
- reliability
- validity
- educational effect

Two of four researchers (LvdE, EP, AK and HM) analyzed each article to determine whether these four aspects were evaluated, which research methods were used and the outcomes of the study. If there was disagreement, a third or fourth researcher also read the article and consensus was reached through discussion. The data are

presented in tables at the end of this chapter (Appendices 1-6). If an instrument was the subject of more than one article, additional articles were only included if they contained new information about the aspects of interest. Based on the tables, the researchers identified highlights and interesting results for each characteristic, which are reported in the results section.

Results

Descriptive analysis

The initial search yielded 349 articles. Of these, 261 were excluded based on the title, a further 50 were excluded based on the abstract and another 19 were eliminated after the reading of the full article. This left a total of 19 articles. The second search yielded 34 articles. After exclusion of five, nine and five articles based on title, abstract and full text, respectively, 15 articles from the second search met the criteria. The manual search of reference list yielded another 5 articles. The resulting 39 articles dealt with 18 different assessment instruments [2-6, 10-43], which are listed in Table 1.

1	Mini clinical evaluation exercise
2	Ophthalmic clinical evaluation exercise
3	Palliative care clinical evaluation exercise
4	Professionalism mini evaluation exercise
5	Competence based assessment, rheumatology
6	Structured clinical observation
7	Patient evaluation assessment form
8	Global rating form in anaesthesiology
9	Ward rating form (in clinical work sampling approach to in-training assessment)
10	Clinical-performance biopsy instrument
11	Clinical evaluation exercise (in emergency medicine training programme)
12	Clinical skills assessment form, direct observation exercise
13	Standardized direct observation assessment tool
14	Evaluation of consulting skills (of trainee general practitioners)
15	Longitudinal evaluation of performance
16	MiniCard
17	Clinical Encounter Card
18	Bedside formative assessment

Table 1 Assessment instruments

Characteristics

The instruments included in the review assess a wide range of competencies or combinations of competencies. Some allow coverage of broad content and can be used in all kinds of clinical situations; others assess content that is limited to a particular setting, e.g. a palliative care or psychiatry clerkship. All instruments itemize content globally, but some are more detailed than others (items such as: 'open-ended questions' versus 'patient communication'). Most items relate to the 'medical expertise', 'communication' and 'professionalism' competencies from the Canadian Medical Educational Directives for Specialists (CanMEDS). Some items relate to the CanMEDS competence 'management skills'. Generally, the instruments appear to be flexible with regard to content. They can be used to assess a multitude of competencies and are easily attuned to a specific educational context.

Most instruments are (intended to be) used for formative purposes. It is consistent with this purpose that almost all instruments ask for qualitative, narrative feedback to be provided in writing or orally. Additionally, almost all instruments require quantitative feedback on a rating scale. These scales vary widely, ranging from dichotomized scores of 'satisfactory' and 'unsatisfactory' to an 11-point scale. A minority of instruments (four with small and three with large scales) provide criteria for the allocation of marks or behavioural anchors. Only one study examines the effects of different rating scales [13] by comparing the results of 9- and 5-point scales. Inter-rater reliability was similar for both scales, but the 9-point scale showed better agreement with previously established levels of competence of a performance on video (the scripted competence level). Based on the assumption that previously established competence levels are accurate, the 9-point scale was better able to accurately classify learners' competence as unsatisfactory or superior. A reference norm for competence rating is specified in no more than eight instruments: five use an 'end of training' norm and three a 'class level' norm. However, norm selection is not based on evidence and authors generally state few or no arguments to support their choice of rating scale or frame of reference, thereby leaving much freedom of interpretation to assessors.

Assessors almost always receive some form of training before an instrument is implemented. Training involves verbal instruction or a workshop, but it is uncommon for training effects to be evaluated. The only authors to do so are Cook et. al.[14], who evaluated the effects of a workshop on error training, performance dimension training, behavioural observation training and frame of reference training using lecture, video and facilitated discussion. They found no improvement in inter-rater reliability of mini-CEX scores in a group of assessors who had attended the workshop compared to a control group receiving no training. Generally, learners instruction receives scant attention. If learners are instructed at all they receive verbal or written instructions, but no studies evaluate the effects.

In conclusion, instruments show considerable variation in content, rating scale, frame of reference, assessor training and learner instruction. There is a striking paucity of research on these characteristics, which are merely described in the majority of studies without evidence to support their value.

Feasibility
Studies of feasibility mostly focus on completion rates of the instruments or users' satisfaction. Feasibility is generally qualified as good but no clear criteria are set in advance and results vary. Durning et al.[18] and Torre et al.[41], for example, report completion rates of 96.4 and 100%, respectively, but Turnbull et al.[42] conclude that feasibility is good with a response rate of only 23%.

Conclusions regarding the feasibility of the various instruments, with the exception of the mini-CEX, are based on single studies. When more studies are available, the results are both negative and positive. Wilkinson et al.[43] attribute feasibility problems to lack of time and the fact that the procedure is experienced as time consuming. Alves de Lima et al.[10] blame poor feasibility on inadequate implementation. They conclude that assessment instruments must be well integrated within the curriculum and part of the routine of practice, and additionally propose that workshops are a better way to implement an instrument than written instructions. Clearly, further studies are needed to unravel the instruments' feasibility issues.

Reliability
Generalizability or reproducibility was studied in four instruments in eight studies. The results are presented in Table 2. We used the Spearman-Brown formula to calculate the average reliability coefficient for all instruments. For most of them acceptable reliability (> 0.8) can be achieved with a sample of 10 encounters. In other words, reliability seems achievable with a feasible sample of encounters. For some studies, we could not determine the number of assessors involved. The study by Margolis et al.[30] is the only one to examine reliability with different numbers of assessors. The results show that one assessor taking 10 encounters is much less reliable than 10 assessors taking one encounter each (0.39 and 0.83, respectively). This outcome is contradicted by Nair et al.[31], who concluded that the mini-CEX is reliable (0.88) with one assessor and eight encounters. However, this study did not explicitly examine the effects of different numbers of assessors. More research is needed to systematically tease out sources of variance in reliability to enable well founded recommendations with regard to the required numbers of (different) assessors and encounters.

Ringsted et al.[37] explain the low inter-rater reliability of their 'global rating form in anaesthesiology' by staff being unfamiliar with the instrument's underlying concept. They suggest that intensive assessor training might improve reliability results, but

the opposite conclusion is put forward by Cook et al.[14]. This conflicting evidence underlines the need for more research into inter-rater reliability and how it is affected by assessor training.

Validity
Criterion validity of the mini-CEX and the 'clinical encounter card' was evaluated by comparison of the results with those of instruments of proven validity. For the mini-CEX, strong and significant correlations were found with results on the Royal College of Physicians and Surgeons of Canada Comprehensive Examination in Internal Medicine (RCSPC-IM), a high-stakes assessment of clinical competence. Correlations were 0.73 with the subscale 'structured oral', 0.67 with the subscale 'bedside station' and 0.72 with the subscale 'written examination'[22]. In addition, strong correlations are reported between mini-CEX scores and corresponding scores on a monthly evaluation form and 'in-training examination scores'[18]. The 'clinical encounter card' showed significant positive correlation with learners' 'clinical performance ratings', 'final grades' and scores on an important summative examination (National Board of Medical Examiners [NBME])[36]. Interestingly, no correlations are reported between the 'clinical encounter card' and an objective structured clinical examination (OSCE). A number of studies infer construct validity from an increase in ratings over time. Kogan et al.[26] report an increase in mean scores on the mini-CEX during one year. Links et al.[28] found significant improvement in skills as manifested in pre- and post-observations, using the 'clinical skills assessment form'. Prescott-Clements et al.[35] report improvement in ratings on 'longitudinal evaluation of performance' in the course of 1 year.

In conclusion, the validity of the mini-CEX and the 'clinical encounter card' appears to be supported by strong and significant correlations with other assessment instruments. For some other instruments positive indications for construct validity are reported, but for most instruments evidence of validity remains to be provided.

Educational effect
Some studies evaluated educational effect by eliciting learners' or assessors' attitudes towards the use of the instrument, but none of the studies examined educational effects by measuring improvement of clinical skills or the quality of patient care. Although authors emphasize the formative nature of assessment procedures, they examine effects on learning and performance by evaluating users' subjective judgments or perceived satisfaction. For the most part, the reported effects are positive. Learners rated the value of 'structured clinical observation' four on a five-point scale[5] and rate the 'clinical skills assessment form' as the second most valuable component of their clerkship in terms of assisting skill acquisition[28]. Outcomes of a student questionnaire on 'bedside formative assessment' show that

Reference number	Instrument	Raters	Encounters	Reliability coefficient	Reliability with 8 encounters (Spearman Brown formula)	Reliability with 10 encounters (Spearman Brown formula)	Reliability with 12 encounters (Spearman Brown formula)
Waas et al. (2001)	Mini-CEX	8	16	≥ 0.80	≥ 0.25	≥ 0.71	≥ 0.75
Margolis et al. (2006)	Mini-CEX	1	10	0.39	0.34	0.39	0.43
Margolis et al. (2006)	Mini-CEX	10	10	0.83	0.57	0.83	0.85
Nair et al. (2008)	Mini-CEX	1	8	0.88	0.88	0.90	0.92
Alves de Lima et al. (2007)	Mini-CEX			10 evaluations for a minimally reliable inference		Reliable	Reliable
Kogan et al. (2003)	Mini-CEX	4	Probably 4	0.62	0.77	0.80	0.83
Kogan et al. (2003)	Mini-CEX	6	Probably 6	0.71	0.77	0.80	0.83
Kogan et al. (2003)	Mini-CEX	8	Probably 8	0.77	0.77	0.81	0.83
Cruess et al. (2006)	P-MEX	Probably 1	10 (a 12)	≥ 0.80	≥ 0.76	≥ 0.80	≥ 0.83
Turnbull et al. (2000)	WRF	Probably 1	3.2 forms completed	≥ 0.70	≥ 0.85	≥ 0.88	≥ 0.90
Richards et al. (2007)	CEC	7	20	0.58	0.36	0.41	0.45
Richards et al. (2007)	CEC	12	18	0.69	0.49	0.55	0.60
Total					0.59	0.69	0.73

Table 2 Generalizability analysis

95,6% recognize its learning value, 70% acknowledge the informative, advisory and motivational role of feedback and 71,9% report that the assessment stimulated them to do more preparatory reading.

However, outcomes like learning behaviour, transfer of skills to new situations or improvement of patient care are not investigated, although they are crucial for the evaluation of educational impact. Currently, educational effects are a neglected area of assessment research, which should be given much greater priority in future research.

Discussion

As for the similarities and differences between the characteristics of the instruments, the main conclusion is that there is huge variation in the competencies being assessed, rating scales, frame of reference, assessor training and learner instruction. Unfortunately, there is hardly any sound research reported on these characteristics. Authors describe rating scales, frames of reference and assessor training but fail to elaborate on rationales and usually do not investigate their value. Consequently,

assessment characteristics remain implicit and interpretation is largely left to assessors. This will inevitably have a profound effect on instruments' measurement characteristics.

Almost all the instruments discussed in this review originated after the introduction of the mini-CEX at the Medical College of Pennsylvania, Philadelphia. An exception is the 'clinical skills assessment form', an observation exercise that was introduced in the psychiatric clerkship at McMaster University, Canada as early as 1984[28], well before the publication of the first paper on the mini-CEX. It is interesting to note that this early appearance on the medical education scene of a predecessor of the mini-CEX apparently failed to make much of an impact either in the literature or in educational practice. Perhaps the time was not ripe then for this type of instrument.

Some information on the feasibility, validity, reliability and educational effect of the instruments we studied emerges from the review. Conclusions regarding feasibility are generally positive. Despite the absence of direct compelling evidence, we are inclined to conclude that training may be the key to effective implementation of instruments because it can improve the quality of their use. The value of these instruments lies mainly in the process of formative feedback and thus in the feedback

skills of assessors and the extent to which they pay serious attention to this process. Much of what is assessed is left implicit and is up to the discretion of assessors[24,33]. Assessors need training to reliably rate learners' performance and discriminate between performance levels[8]. For learners too training may play an important role, although no direct evidence is available to support this. It seems likely that learner training can increase feasibility and educational effect.

Criterion validity was only evaluated for the mini-CEX and the 'clinical encounter card', and these instruments showed strong and significant correlations with other assessment instruments. Construct validity was inferred from three studies showing that ratings increased over time. Otherwise, like Kogan et al.'s review of validity[8], our review reveals a general lack of evidence of validity.

The outcomes of reliability studies suggest that around 10 encounters suffice for a reproducible outcome. This is somewhat surprising. In terms of testing time (time of one medical consultation) 10 encounters compares favourably with the samples needed for other standardized and objectified assessment formats[44], although one would expect poor reliability of an instrument characterized by absence of explicit characteristics. Apparently (different) assessors pick up measurement information that is relatively generalizable across individual encounters, while at the same time broad sampling across assessors evens out the effects of assessor subjectivity. Good reliability is no guarantee for the absence of bias, however, and, due to their quite subjective nature, instruments like the mini-CEX may actually be quite vulnerable to bias. All this requires further investigation. We also need more evidence regarding factors that contribute to (un)reliability and the extent of this contribution to underpin recommendations on sound sampling strategies.

Evidence on educational effect is lacking as well. No studies examined whether instruments improve learning, clinical skills or the quality of patient care. Given the formative nature of the instruments, effects on learning and performance are more or less the prime objective of this type of assessment. Existing research typically evaluates perceptions of users, and although the outcomes are overwhelmingly positive, they do not provide compelling evidence for learning effects. More rigorous research will have to elucidate the educational effects of clinical work-based assessment.

An important conclusion from our review appears to be that instruments for authentic work-based assessment of single clinical encounters should not be evaluated outside the context of the curriculum or other assessment instruments. Assessment by one instrument can only be a part of the whole story. The 'competence based assessment, rheumatology' for example was not valid when applied in isolation[17]. It should be used as a component of a spectrum of assessment instruments that complement each other. While optimization of the feasibility, validity, reliability and educational effect of individual instruments is important, it is equally, if not

more, important to look from a broader perspective at the respective unique contributions of different instruments to the assessment of clinical competence[44]. Assessment procedures should be integrated within the curriculum and preferably also be an integral part of routine practice[10].

It should be noted that we included articles in the review on the basis of the subjects they addressed, not the quality of their research. Some bias may have arisen because we did not systematically judge research quality.

In so far as the articles report on feasibility, validity, reliability or educational effect, the conclusions are mostly positive. This absence of negative or critical outcomes could be suggestive of publication bias. It cannot be ruled out that studies on inadequate instruments were not published.

Although single-encounter clinical assessment instruments appear to be received positively in the literature, this positive reception is based on relatively limited empirical justification. Results on the most extensively evaluated aspects, feasibility and reliability, support the viability of the format and the use of a minimum of 10 encounters to attain reliability. However, there is an obvious need for further, and especially more scientifically rigorous, research on all the characteristics that we studied. We also need further research on basic characteristics like rating scales, narrative feedback, frame of reference, etc. Although a call for more and better research may be the sad conclusion from most reviews, it is unfortunately equally applicable to single-encounter work-based clinical instruments.

	Setting	Year of (first) publication	Formative/ summative	Trainee (T) and rater (R)	Encounters	Competences	Descriptions of behaviours?	Scale	Criteria for allocation of marks	Frame of reference
Mini - CEX	Internal medicine residency, Philadelphia, USA. International medical graduates of Australia Also used in USA, Canada, UK, Argentina.	1995	Both	T: resident R: attending, senior resident	Inpatient, outpatient, ED, other (including admission, discharge)	1. medical interviewing 2. physical examination 3. humanistic qualities/ professionalism 4. clinical judgement 5. counseling 6. organization/ efficiency 7. overall clinical competence	Yes, (total of 26, specified per competency) (www.abim.org/ pdf/paper-tools/ minicex.pdf)	9-point scale: 1-3 = unsatisfactory 4-6 = marginal/ satisfactory 7-9 = superior Plus: 'not observed'.	Specified by ABIM for resident assessment. Unsatisfactory= consistently falls short of reasonable expectations. Marginal= in general meets some expectations but occasionally falls short. Satisfactory= always meets and occasionally exceeds reasonable expectations. Superior= far exceeds reasonable expectations.	Only 4 on 9-point scale is specified as: marginal; conveys expectation that with remediation resident will meet certification standards. Otherwise not specified.
Mini- CEX in clerkship	(Internal) medicine clerkship, Canada, USA.	2002	Both	T: student R: attending, resident	Inpatient and outpatient	1. medical interviewing 2. physical examination 3. humanistic qualities/ professionalism 4. clinical judgement 5. counseling 6. organization/ efficiency 7. overall clinical competence	No (not mentioned)	9-point scale: 1-3 = unsatisfactory 4-6 = marginal/ satisfactory 7-9 = superior Plus: 'not observed'.	Not specified	Not specified
OCEX	Ophthalmology residency, USA	2004	Formative	T: resident R: attending	New patients (n= 3-4 per year)	Assessment of: -patient care -professionalism -interpersonal skills -medical knowledge Categories on form: 1. interview skills 2. physical exam 3. interpsonal skills/ professionalism 4. case presentation	Description of 33 sub-items, some are behaviours and some regard specification of the category-subject.	4-point scale: 1= does not meet expectations 2= meets some expectations 3= meets all expectations 4= exceeds expectations Plus: 'not applicable'.	Allocation of score 1 to 4 is specified for each sub-item in a scoring rubric (*Golnik et al., 2004, table 1*).	Not specified
Palliative care CEX	Internal medicine residency, university of Pittsburgh, USA	2005	Formative	T: (1st year) resident R: one of the clinicians of 'palliative care service' (physicians, nurse, medical ethicist and social worker	Giving bad news to or discussing wishes regarding 'code status' with seriously ill patient (n=1)	Communication skills in giving bad news and discussing code status	Yes (total of 18, see: *Han et al., 2005, table 3*)	Not specified	Not specified	Not specified

Appendix 1 Characteristics

	Setting	Year of (first) publication	Formative/ summative	Trainee (T) and rater (R)	Encounters	Competences	Descriptions of behaviours?	Scale	Criteria for allocation of marks	Frame of reference
P-MEX	Clerkship in: internal medicine, general surgery, paediatrics, psychiatry and obstetrics/ gynaecology, McGill university, Montreal, Canada	2006	Both	T: 3rd- 4th year student R: faculty and senior resident (multiple raters per student, number not specified)	Situations where student's behaviour can be observed (e.g. patient encounters, small group sessions, sign-out rounds)	Professional behaviours: 1. doctor-patient relationship skills 2. reflective skills 3. time management 4. interprofessional relationship skills	Yes (total of 24, specified per competency, see: *Cruess et al., 2006, table 1*)	4-point scale: 1= unacceptable 2= below expectations 3= met expectations 4= exceeded expectations Plus: 'not observed or not applicable'	Not specified	Not specified
CBA- rheuma- tology	Rheumatology	2006	Formative	T: resident R: consultant in rheumatology (n=1-2 per trainee)	New and follow-up outpatients with variety of rheumatology problems (n= 4-7 in a 3-hour clinic)	1. history taking 2. examination 3. diagnostic skills 4. management plan 5. communicative skills 6. letter dictation 7. overall impression	No	3-point scale: 1= does not reach standard 2= borderline 3= good	Allocation of score 1, 2 and 3 are specified *(Dowson & Hassell, 2006, table 3).*	Not specified
SCO	Paediatric clerkship, Jefferson Medical College, Philadelphia, USA	2000	Formative	T: 3rd year student R: general paediatrician (multiple raters, number not specified)	Part of paediatric patient encounters (max. 3 min of observation, not specified whether inpatient or outpatient)	- history taking or - physical examination or - information giving	Yes (total of 47, specified per competence) -history taking: total of 22 items -physical exam: total of 8 items -information giving: total of 21 items *(see: Lane & Gottlieb, 2000, table 1,2,3)*	Not specified	Not specified	Not specified

Appendix 1 Continued

Appendix 1 Continued

	Setting	Year of (first) publication	Formative/ summative	Trainee (T) and rater (R)	Encounters	Competences	Descriptions of behaviours?	Scale	Criteria for allocation of marks	Frame of reference
PEAF	Surgical residency (general surgery and trauma rotations), Michigan State university, USA	2005	Formative	T: PGY I and II resident R: faculty	New outpatient (scheduled during teaching patient hour)	1. punctuality/ introduction/ describes role 2. quality/ appropriateness of history taking skills 3. open ended questions/ listening/ interviewing skills 4. quality of physical exam/ attention to patient autonomy 5. knowledge of test results/ images 6. ability to formulate and discuss assessment/ classic presentation/ differential diagnosis 7. knowledge of evidence-based data 8. decision making skills/ knowledge of steps, orders/ plan 9. ability to educate others/ communicate with team 10. quality/ timeliness of data entry/ dictation Plus: 'overall score for level of training.	No	For each item a score on a 11-point scale: 0-100%, increments of 10% And a 8 letter scale for 'overall score for level of training': A+ = best resident A = superb B+ = exceeds expectations B = solid C+ = below expectations C = marginal D = poor performance F = failure	Not specified	For the 11-point scale: novice, beginner, advanced, expert, mastery PGYI is expected to perform at novice or beginner level (0-30%); PGYII at advanced level; senior residents at 80-100% 8-letter scale: score compares resident to other residents of same level.
GRFA	Anaesthesio-logy residency, Denmark	2003	Summative *(NB, used in research setting)*	T: resident R: consultant in anaesthe-siology	-insertion of epidural catheter -pre-op consult -emergency induction of anaesthesia -round on patient in ICU *(NB, research purposes, not described for which purposes it was intended in practice)*	1. patient communication 2. clinical and technical skills 3. knowledge	No	5-point scale: 1= clear fail 2= borderline fail 3= borderline pass 4= clear pass 5= excellent Plus: an overall dichotomous scoring of pass/fail	Not specified	Mastery (all elements of performance must be correct)

	Setting	Year of (first) publication	Formative/ summative	Trainee (T) and rater (R)	Encounters	Competences	Descriptions of behaviours?	Scale	Criteria for allocation of marks	Frame of reference
WRF	Internal medicine clerkship, university of Ottawa, Canada	2000	Potentially formative	T: 3rd year student R: attending	Inpatient discharge	1.therapeutic strategies 2.communication skills 3.consultation skills 4.management skills 5. interpersonal behaviours 6. continued learning skills 7. health advocacy skills Plus: a global rating of overall performance	No	5-point scale: 1= unsatisfactory 2=? 3=? 4=? 5= excellent	Not specified	Not specified
CPB	Family practice residency, USA	2002	Not specified	T: resident R: attending	Outpatients in family practice	1.history taking and physical examination 2.interpersonal skills 3.clinical problem solving	No	Likert-scale, not further specified	Not specified	Not specified
CEE	Emergency medicine residency, Atlanta, USA	2002	Formative	T:resident R:attending	'typical' patient encounter in ED	Categories on form: -emergency stabilization -data-gathering-history -data-gathering- physical examination -case presentation -diagnosis and plan for diagnostic studies and medical care According to authors the following competences are assessed: -patient care -practice-based learning -interpersonal and communication skills -professionalism -system-based practices	Yes, total of 27, specified per category. Items are scored, not category as a whole	3-point scale: 1= below expected 2= meets expected 3= outstanding	1= falls short of reasonable expectations 2= always meets and occasionally exceeds expectations 3= far exceeds expectations for level of training	Level of training
CSAF	Psychiatric clerkship, McMastery university, Hamilton, Canada	1984	Formative	T: student R: clinical supervisor	New psychiatric patients (inpatient or outpatient not specified)	-interviewing -history taking -mental status examination -doctor-patient-relationship -problem formulation -treatment plan	Yes, total of 17, specified per competence (see: Links et al., 1984, table 1)	7-point scale not further specified	1= non or few features were demonstrated 2-6 = not specified 7= criterion performance	Criterion or ideal performance

Appendix 1 Continued

	Setting	Year of (first) publication	Formative/ summative	Trainee (T) and rater (R)	Encounters	Competences	Descriptions of behaviours?	Scale	Criteria for allocation of marks	Frame of reference
SDOT	Emergency medicine residency, USA	2006	Not specified	T: resident R: attending	Patient encounters in emergency department	1st assessment on form: 4 categories; -data fathering -synthesis/ diagnosis -management -disposition 2nd assessment on form 5 competencies: -patient care -medical knowledge -interpersonal and communication skills -professionalism -system-based practice Plus: 'overall clinical competence'	Yes, total of 26, specified per competence. These 26 items are separately scored	1st assessment and overall clinical competence: 3-point scale: 1= needs improvement 2= meets expectations 3= above expected Plus: 'not assessed' 2nd assessment 5-point scale: Score 1-5 in combination with -needs improvement, -meets expectations, -above expectations *(NB, it is not completely clear which score combines with which description)*	Extensive (7-page) description of behavioural anchors for each item (www. cordtests.org)	Level of training
ECS	General practice residency, West Australia	1997	Formative	T: resident R: experienced GP	New and follow-up patients in general practice	Communication skills: 14 behaviours are to be scored *(Nyman & Sheridan, 1997, figure 1)*	No further specification	5-point scale: 1= strongly disagree 2= disagree 3= unsure 4= agree 5= strongly agree Plus: 'not relevant'	Not specified	Not specified
LEP	Dentist, dental vocational training, Scotland	2002	Formative (can contribute towards a summative decision)	T: post-graduate dental trainee R: clinicians, advisers, nurses, other members of the clinical team	Almost any clinical situation across dentistry	1.examination and consultation skills 2. clinical judgement and diagnosis 3.technical ability and manual dexterity 4. communication skills 5. professionalism 6. knowledge 7. organisation	No further specification	9-point scale: 1-3 = need improvement 4-6 = satisfactory 7-9 = superior	Not specified	Completion of training

Appendix 1 Continued

In-training assessment using direct observation of single-patient encounters: a literature review

In-training assessment using direct observation of single-patient encounters: a literature review

	Setting	Year of (first) publication	Formative/ summative	Trainee (T) and rater (R)	Encounters	Competences	Descriptions of behaviours?	Scale	Criteria for allocation of marks	Frame of reference
MiniCard	Internal medicine, 4 different residency programmes in the USA	2008	Probably both	T: resident R: internal medicine faculty member	Problem commonly seen in internal medicine	Four sections: 1.history 2.physical 3.presentation of plan 4.counselling Each section consists of three domains: a. interpersonal communication b. medical knowledge c. professional domains	3 to 11 prompts given for each domain	4-point scale: 1= excellent 2= good 3= marginal 4= poor	Each rating category has adjectival and behavioural anchors specific to the domain	Level of training for medical knowledge, mastery level for interpersonal communication and professionalism
CEC	A: surgery clerkship, USA B: internal medicine clerkship, Canada	1999	A + B: formative	A + B: T: 3th year medical student R: faculty or PGY resident	A + B: inpatient and outpatient, new and follow-up	A + B: 1.history taking 2.physical examination 3.professional behaviour 4.technical skills 5.case presentation 6.problem formulation (diagnosis) 7.problem formulation (therapy) 8.other	A + B: no further specification	A: 6-point scale: 1 = unsatisfactory 2 = below level of average 3 = at level of average 4 = above level of average 5 = outstanding above level of average 6 = at level of intern B: 5-point scale: 1= unsatisfactory 2= below level of average 3= at level of average 4= above level of average 5= at level of intern	A + B: not specified	A + B: norm-referenced (not further specified)
BFA	Medicine clerkship at the university of Cape Town, South Africa	2006	Formative	T: 4th year medical student R: clinician educators, all specialist physicians with 5 years teaching experience	'Blinded' patient encounters: student examines inpatient without access to their clinical records	1.basic clinical skills a. information b. report 2.diagnostic reasoning skills 3.knowledge a. investigations b. plan	Descriptions given on the form	9-point scale: 1-3 = poor 4-6 = adequate 7-9 = good	Not specified	Not specified

Appendix 1 Continued

	Type of feedback	Additional information on feedback	Way of training raters	Instruction of trainees	Form displayed in article?	Instrument retrievable on internet?
Mini-CEX	Quantitative: yes Qualitative: - verbal: yes - written: yes (space on form for comments)	- 80% of sessions lead to recommendations for improvement - in 61% learner interaction was promoted - in 34% self-assessment occurred - in 10% action plans were developed (in 3 out of 11 sessions this was generated by the resident)	Raters were instructed on goals of mini-CEX (including importance of feedback to promote improvement of clinical skills) and an instructional ABIM-handout was provided. - In the study of Cook et al., 2008 intervention of an intensive workshop was studied. The workshop included rater error training, performance dimension training, behavioural observation training, frame of reference training using lecture, video and facilitated discussion. Inter-rater reliability was higher after the workshop than before. - *In other studies on mini-CEX: no training or not specified.* - *A written instruction for raters can be found on the ABIM homepage (www.abim.org/pdf/paper-tools/attending.pdf)*	- instruction for trainees is not specified in articles except for the research on the workshop in the article of Cook et al., 2008. - *A written instruction for trainees can be found on the ABIM homepage (www.abim.org/pdf/paper-tools/resident.pdf)*	Yes	Yes: guideline, form, written instruction for raters and trainees. http://www.abim.org/pdf/paper-tools/minicex.pdf http://www.abim.org/pdf/paper-tools/attending.pdf http://www.abim.org/pdf/paper-tools/resident.pdf
Mini-CEX in clerkship	Quantitative: yes Qualitative: - verbal: not specified (probably after observation) - written: not specified	-	Raters were oriented to rationale and use of mini-CEX during a meeting. Outpatient faculty were also mailed an orientation package. There was no formal training to use the form	Instruction was limited to number of forms to be handed in.	No *(Mini-CEX form is used with some adjustments)*	No
PDA-based	Quantitative: yes Qualitative: - verbal: not specified - written: yes	Feedback was provided in 96%	No formal training. Before start of rotation, raters received orientation package and e-mail in which purpose of mini-CEX was explained, key features of the PDA-based form were discussed and instructions were provided on how to structure the exercise and complete the form. There was no detailed guidance on how to deliver feedback.	Trainees were oriented to form and procedure by clerkship director during clerkship orientation session. Purpose of form and rationale for each item was reviewed in detail. Trainees were instructed to assist raters in the use of the PDA-form if necessary.	Yes	
OCEX	Quantitative: yes Qualitative: - verbal: probably, but not specified in the article - written: yes (space on the form for comments)	Resident is to receive immediate feedback.	Raters received a form, scoring rubric and a 3 minute explanatory introduction video-CD. (*NB, research purposes*)	Not specified (*studies did not regard resident use of OCEX*)	Yes	Http://www.acgme.org/acWebsite/RRC_240/240_OCEX.pdf
Palliative care CEX	Quantitative: probably, however not specified Qualitative: -verbal: yes -written: not specified	Feedback included a review of both residents' and observers' impressions of the strengths and weaknesses of the discussion with the patient	Raters were individually trained (by authors) on all aspects of the study protocol, procedures for conducting the palliative care CEX and proper scoring of the instrument to encourage consistency and reliability.	Raters provided a review of the circumstances, goals and strategies for the discussion with the patient.	No	No

Appendix 2 Characteristics

	Type of feedback	Additional information on feedback	Way of training raters	Instruction of trainees	Form displayed in article?	Instrument retrievable on internet?
P-MEX	Quantitative: yes Qualitative: -verbal: not specified -written: yes (space on the form for comments)	Rater is expected to give timely feedback to student	Instructions on using the form were given to all raters. Not further specified	Not specified	No	No *Only a presentation on NVMO-website with example of P-MEX:* *http://www.nvmo.nl/portals/0/congres200/Steinert.pdf*
CBA rheuma-tology	Quantitative: yes Qualitative: -verbal: not specifically mentioned, however 'constructive feedback was given during and/or at the end of the assessment' -written: yes (space on the form for comments on each item); 1 form for each encounter in the session and 1 summary form at the end of the session	-	Raters were given information regarding aims, format and scoring criteria for the assessment. Scoring criteria were shown at least one month before assessment. Not further specified, no training sessions	Trainees were given information regarding aims, format and scoring criteria for the assessment. Scoring criteria were shown at least one month before assessment. Not further specified, no training sessions.	Yes	No
SCO	Quantitative: no Qualitative: -verbal: yes -written: ?	Observation was followed by immediate feedback, outside patient room, lasting maximally 2 minutes. The short observation would limit the number of feedback points, so that feedback did not become overwhelming for the student and would not take more than a few minutes for the rater to complete. Since there were repeated observations, students would be given a chance to incorporate what they had been taught, feel that they were gaining mastery and improving their clinical skills. Raters were instructed to write down exactly what trainees said and did, to give focussed and effective feedback. If 1 or 2 major feedback points are identified in the first few seconds of observation, the SCO is ended.	Two-hour workshop, consisting of: -discussion of rationale for SCO program -overview of the SCO methodology -overview of basic tenets of effective feedback -discussion of procedures -practice in observing and using SCO forms by looking at 2-3-minute video vignettes of students interacting with patients -role play to practise giving feedback	Students were oriented to SCO program at start of 6-week clerkship and again just before starting the SCO program (a 2-week block). The skill guidelines were reviewed and discussed, procedures were discussed and it was emphasized that it was a teaching experience not an evaluation.	No	No

Appendix 2 Continued

	Type of feedback	Additional information on feedback	Way of training raters	Instruction of trainees	Form displayed in article?	Instrument retrievable on internet?
PEAF	Quantitative: yes Qualitative: -verbal: ?: 'comments were invited on each case reviewed'. -written: yes, space on form for comments/ improvements	-	Raters were trained on use of the evaluation form. Not further specified	Not specified	Yes	No
GRFA	Quantitative: yes (in wording) Qualitative: -verbal: no -written: ?: participants were invited to give written comments, not specified if this included feedback, or only comments about the instrument.	Participants were invited to give written comments, not specified if this included feedback, or only comments about the instrument. The instrument was used in a study situation with video-taped encounters.	The instrument was used in a study situation. Raters were instructed to: -use their professional expertise and judgement in evaluation of resident's performance -make an overall decision regarding pass or fail according to assessment protocol (criterion for passing was mastery)	Not applicable because of the study situation where a standardized resident was rated.	Yes	No
WRF	Quantitative: yes Qualitative: -verbal: yes -written: yes	Formative feedback on each performance was accommodated by a comment section on the form.	Two-hour workshop, followed by monthly communication reminding raters of the basis of the projects and the need to complete evaluations.	Students were oriented at beginning of rotation and subsequently met with research associate on weekly basis to discuss issues or questions related to the project.	No	No
CPB	Quantitative: yes Qualitative: -verbal: ? -written: yes	Specificity of feedback improved significantly after intervention as described in the next cell, (pre-intervention score mean of 5.25/9, post-intervention mean score 7.44/9; p<.01). Timelines and immediacy of feedback also improved dramatically (4.43/9 pre-intervention, 7.81/9 post-intervention, p<.01.	2-hour course in which; -basic outline of evaluation was reviewed -criteria for the standardized assessment and completion of the CPB were discussed. The format of the course contained the analysis of 5- to 10-minute videotaped vignettes of outstanding, satisfactory and unsatisfactory resident performances.	Not mentioned	No	No
CEE	Quantitative: yes (in wording) Qualitative: -verbal: yes -written: yes (space on form for comments on each category)	Rater is to provide immediate feedback on encounter in oral and written form.	No / not mentioned	No / not mentioned	Yes	No
CSAF	Quantitative: yes Qualitative: -verbal: yes -written: ?	Following the encounter, there was a 30-min feedback session, during which the student presented a problem-formulation and management plan.	No / not mentioned	No/ not mentioned	No ('further details about the CSAF can be obtained from the first author')	No

Appendix 2 Continued

In-training assessment using direct observation of single-patient encounters: a literature review

Chapter 2

	Type of feedback	Additional information on feedback	Way of training raters	Instruction of trainees	Form displayed in article?	Instrument retrievable on internet?
SDOT	Quantitative: yes Qualitative: -verbal: ? -written: yes, (space on form for summary comments and comments for each competence)	-	For reliability study: raters were briefed on the purpose of SDOT.	Not specified for practice (not applicable for reliability study, because standardized residents were used).	No, although authors refer to website where SDOT is available	Yes: www.cordtests.org
ECS	Quantitative: yes Qualitative: -verbal: ? -written: no	-	Detailed set of written guidelines was provided for raters	Not mentioned	Yes	No
LEP	Quantitative: yes Qualitative: -verbal: ? -written: yes	-	All participants were trained (not further specified)	All participants were trained (not further specified)	Yes	Trainers guide http://www.nes.scot.nhs.uk/ dentistry/dvt_htvt/assessment/ trainers/documents/LEPA trainersguide1.pdf
MiniCard	Quantitative: yes Qualitative: -verbal: yes -written: yes	Qualitative feedback is written in the research design, but given verbal in real-world situations.	1-hour training session with videotaped description of the tool, followed by a practice session.	Not applicable (study situation where a standardized resident was rated)	Yes	No
CEC	Quantitative: yes Qualitative: -verbal: yes -written: yes	-	The encounter cards were presented at a number of forums, including departmental meetings, academic half-day for internal medicine residents and an educational faculty retreat *(Hatala & Norman, 1999)*. Evaluators were instructed to rate performance on a global scale and give verbal and written feedback. A letter explained the system *(Paukert et al., 2002)*.	Students were informed and CEC's were explained (not further specified).	Yes	No
BFA	Quantitative: yes Qualitative: -verbal: probably -written: yes	-	Clinician educators who volunteered to participate in the study attended 2 workshops in which the principles and purpose of formative assessment were discussed and the structured feedback form was designed.	Not mentioned	Yes	No

Abbreviations used in Appendix 1 and 2: **ABIM** = American Board of Internal Medicine;

ED = Emergency Department; **PGY** = Postgraduate year; **pre-op** = preoperative; **ICU** = intensive care unit.

Appendix 2 Continued

	Feasibility determined?	How was this performed?	Outcomes
Mini-CEX	Yes	Completion rate *(Alves de Lima et al., 2007; Durning et al., 2002; Wilkinson et al., 2008).* Survey *(Nair et al., 2008).* Focus group *(Malhorta, 2008).*	Low completion rate highlights feasibility problems; lack of time was the main factor preventing completion *(Wilkinson et al., 2008).* Completion rate of 96.43% *(Durning et al., 2002).* 15% of residents were evaluated four or more times during the study period. Feasibility was defined by a minimum of 50% of the residents obtained at least four mini-CEXs. Feasibility in this study was poor *(Alves de Lima et al., 2007).* Never or only occasionally experienced difficulties. Most examiners (14/18) were satisfied or very satisfied *(Nair et al., 2008).* The assessment first reflects anxiety. Repeated mini-CEX experiences made the resident learner more confident in controlling and enforcing the positive educational effect *(Malhorta, 2008).*
Mini-CEX in clerkship	Yes	Completion rate *(Kogan et al., 2002; Kogan et al., 2003).* Satisfaction ratings *(Kogan et al., 2002).*	Completion rate: mean of 7.3 mini-CEX, 9 were asked *(Kogan et al., 2002)* Completion rate: 89% of target *(Kogan et al., 2003).* Satisfaction rate: faculty/ residents: 7.2 Students: 6.8 *(Kogan et al., 2002).*
PDA-based mini-CEX	Yes	Completion rate *(Torre et al., 2007).* Satisfaction rating *(Torre et al., 2007).*	Completion rate: 100% = 354 forms -feedback provided in 96% Satisfaction rate: Student satisfaction 8.0; Resident satisfaction 8.1; Faculty satisfaction 7.4 Researchers conclude that the PDA-based mini-CEX is feasible.
OCEX	No	-	-
Palliative care CEX	Yes	Resident questionnaires (telephone survey and written questionnaire) *(Han et al., 2005).*	Most residents rated the Palliative Care CEX very highly (> 4 on five-point scale) among several dimensions: educational value, overall quality of the experience, preceptor's effectiveness in creating a positive learning experience, improvement in comfort with discussion, importance of formal instruction in end-of-life communication, value of preceptor feedback). The study provides support for the feasibility in this setting.
P-MEX	Yes	Demographic and contextual data were computed *(Cruess et al., 2006).*	-221 forms, 74 students, 47 evaluators -context of evaluation varied greatly Researchers conclude that this study suggests a feasible format for evaluating professionalism in clerkship training.
CBA rheumatology	Yes	Consultant and resident questionnaires *(Dowson & Hassell, 2006).*	12/13 of consultants and 12/12 residents were positive about the process. Researchers conclude that this is a worthwhile exercise for evaluating important clinical and consultation skills.
SCO	Yes	Student and faculty questionnaires *(Lane & Gottlieb, 2000).* Observation rate *(Lane & Gottlieb, 2000).*	The SCO program was highly regarded as a clinical skills teaching tool by both students and faculty. It maintains the number of observations. Researchers conclude that the program was feasible.
PEAF	No	-	-

Appendix 3 Feasibility

	Feasibility determined?	How was this performed?	Outcomes
GRFA	Yes	Participants indicate the appropriateness of the scoring form they had used on a 5-point scale and with written comments *(Ringsted et al., 2003)*.	Clinicians found checklists with task-specific items more appropriate than rating forms with general dimensions of competence.
WRF	Yes	Completion rate *(Turnbull et al., 2000)*.	Mean; 8.1 WRFs submitted per student 23% response rate. While response rates were low, the minimum number needed to achieve adequate level of reliability is exceeded. Thereby meeting requirement for feasibility.
CPB	Yes	Questionnaires – satisfaction levels of residents and receptors *(Ross, 2002)*.	Higher rate of documented direct observation and feedback than previously reported. The specificity of feedback improved significantly (5.25/9 – 7.44/9; p<.01). Timeliness and immediacy of feedback improved dramatically 94.43/9 – 7.81/9; p<.01). Residents and faculty were highly supportive of continuing CPB as part of the evaluation during continuity clinic.
CEE	No	-	-
CSAF	Yes	Compliance of students and supervisors *(Links et al., 1984)*.	Students and staff complied with the exercise. Particularly when the students were given more responsibility for completion of the exercise.
SDOT	No	-	-
ECS	No	-	-
LEP	Yes	Average time spent on each assessment *(Prescott-Clements et al., 2008)*. Rater questionnaires *(Prescott-Clements et al., 2008)*.	Average time spent is 22 minutes for observation and 8 minutes for giving feedback. 65% found implementation easy (only 11% disagreed on this), 63% thinks the assessment had sufficient flexibility.
MiniCard	No	-	-
CEC	Yes	Return rate *(Hatala & Norman, 1999)*.	64% participated, mean of 7.9 encounters with CEC during rotation.
BFA	Yes	Staff questionnaires *(Burch et al., 2006)*.	Most participating clinical educators agreed that the assessment could be satisfactorily integrated into bedside teaching sessions.

Appendix 3 Continued

	Reliability determined?	Type(s) of reliability	How was this performed?	Outcomes
Mini-CEX	Yes	(1) Generalizability *(Alves de Lima et al., 2007; Margolis et al., 2006; Nair et al., 2008; Wilkinson et al., 2008).* (2) Reproducibility *(Norcini et al., 1995).* (3) Internal consistency *(Durning et al., 2002; Hatala et al., 2006).* (4) Inter-encounter reliability *(Hatala et al., 2006).* (5) Inter-rater reliability *(Cook & Beckman, 2008).*	(1) Generalizability analysis *(Margolis et al., 2006; Nair et al., 2008; Wilkinson et al., 2008)* or SEM benchmark and D-coefficient *(Alves de Lima et al., 2007).* (2) Reproducibility coefficients (ratios of universe score variance to total variance) (3) Cronbach's α (4) Cronbach's α (5) ICC	(1) Eight different assessors observing at least two encounters *(Wilkinson et al., 2008).* One rater 10 cases, reproducibility .39. Ten raters 1 case, reproducibility .83 *(Margolis et al., 2006).* G-coefficient for eight encounters was 0.88, suggesting reliability of .90 for 10 encounters *(Nair et al., 2008).* Using the SEM benchmark of .26, a minimum of 10 evaluations was necessary to produce a minimal reliable inference. This corresponds to a D-coefficient of .44 *(Alves de Lima et al., 2007)* (2) Between 12 and 14 encounters are required to reach a reproducibility of .80 *(Norcini et al., 1995).* (3) α = .90 for 162 forms *(Durning et al., 2022)* and α = .79 *(Hatala et al., 2006).* (4) inter-encounter reliability α = .74 *(Hatala et al., 2006).* (5) inter-rater reliability on five-point scale (research format) was .40, on nine-point scale (original mini-CEX) .43. The inter-rater reliability of mini-CEX was similar for 5-point and 9-point scale *(Cook & Beckman, 2008).* Inter-rater reliability was higher after intensive workshop intervention than before the intervention, (ICC .34 vs. ICC .18) but standard error of measurement was similar for both periods *(Cook et al., 2008).*
Mini-CEX in clerkship	Yes	Generalizability *(Kogan et al., 2003).*	Generalizability analysis	Reproducibility for 4 raters is .62, for 6 raters, .71, for 8 raters .77.
OCEX	Yes	(1) internal consistency *(Golnik & Goldenhar, 2005).* (2) inter-rater reliability *(Golnik & Goldenhar, 2005).*	(1) Cronbach's α (2) Analyzing the distribution of ratings across each category for each item. With an acceptable distribution of having at least 85% of ratings occur in two consecutive categories.	(1) Coefficient α for OCEX as a whole = .81; interview skills = .65; examination = .27; interpersonal skills/ professionalism = .73; ease presentation = .70. (2) 94% of the items has at least 85% of the ratings occur in 2 consecutive categories.
Palliative care CEX	No	-	-	-
P-MEX	Yes	Generalizability *(Cruess et al., 2006).*	Generalizability analysis (GENOVA)	Between 10 and 12 encounters are required to obtain reproducibility coefficient of .80.
CBA rheumatology	No	-	-	-
SCO	No	-	-	-
PEAF	No	-	-	-
GRFA	Yes	Inter-rater reliability *(Ringsted et al., 2003).*	Agreement of pass/fail decision	Poor agreement among clinicians regarding pass/fail decisions. Reliability could be enhanced by having several different assessors and several observations, familiarity with the concept, to train the clinicians as assessors.

Appendix 4 Reliability

	Reliability determined?	Type(s) of reliability	How was this performed?	Outcomes
WRF	Yes	Within method reliability/ generalizability (to determine the number of assessments necessary) *(Turnbull et al., 2000).*	Spearman brown prophecy formula	Spearman brown = 3.2 forms necessary to achieve within method reliability (reliability of .70).
CPB	No	-	-	-
CEE	No	-	-	-
CSAF	Yes	(1) Inter-rater reliability *(Links et al., 1984).* (2) Test- retest reliability *(Links et al., 1984).*	(1)Review of a single videotape of a student performance by 12 clinical supervisors. (Satisfactory agreement: agreement plus or minus one on the 7-point scale by greater than 60% of observers). Number of actual agreement over the number of possible agreements. (2)Correlation coefficient between the pre-exercise and post-exercise (5-week interval).	(1)Percentage of agreement varies between 52% and 89%. Overall percentage was 71%. In 13 of 17 cases agreement met arbitrary standard of 60% (2) Significant correlation in 16 of 17 categories (sig. < .05) (N=38).
SDOT	Yes	(1) Inter-rater reliability *(Shayne et al., 2006).* (2) Internal consistency *(Shayne et al., 2006).*	(1) Video presentations of 2 simulated resident-patient encounters. Video 1: average resident performance. Video 2: several significant weaknesses. They (N=82) watched alone and could watch twice before completing the form. Inter-rater reliability on each item and of each combined competency score. (2) Cronbach's α is calculated and intraclass correlation coefficients.	(1) Four of the 26 items were less reliable. Overall reliability may be improved by modifying these four portions. Multivariate analysis found no differences in rating by faculty when examined by experience, academic title, site, of previous use of the SDOT. (2) α = .93. If four items were removed α = .95 ICC = .74. If four items were removed ICC = .81 SDOT appears to be reliable according to the researchers.
ECS	No	-	-	-
LEP	No	-	-	-
MiniCard	Yes	Inter-rater reliability *(Donato et al., 2008).*	Fleiss' kappa	K = .520 – moderate agreement
CEC	Yes	(1) Generalizability *(Richards et al., 2007).* (2) Inter-encounter reliability *(Hatala & Norman, 1999)*	(1) ICC (2) Intraclass correlations (ICC)	(1) ICC .58 on 20 encounters, 7 evaluators. ICC .69 on 18 encounters, 12 evaluators. (2) Inter-encounter reliability of .32 for single encounter. Reliability of .79 overall for 8 encounters per student.
BFA	No	-	-	-

Appendix 4 Continued

	Validity determined?	Type(s) of validity	How was this performed?	Outcomes
Mini-CEX	Yes	(1) Criterion validity *(Hatala et al., 2006).* (2) Criterion validity *(Durning et al., 2002).* (3) Construct validity *(Holmboe et al., 2003).* (4) Construct validity *(Alves de Lima et al., 2007).* (5) Test of halo *(Cook et al., 2008).* (6) Face validity *(Wilkinson et al., 2008).* (7) Accuracy *(Cook & Beckman, 2008).*	(1) Pearson product-moment correlation between mini-CEX and RCSPC-IM. (2) Comparison mini-CEX mean scores with widely used methods for evaluation of residents' clinical competence. Analysis of variance (ANOVA). (3) Variance analysis with ANOVA (video analysis on three clinical skills on three different performance levels). (4) Ability of the mini-CEX to discriminate between pre-existing levels of expertise. Significant mean differences between different years of training. (5) Test of halo. (6) Questionnaires. (7) Scores compared to scripted competence level.	(1) Correlations: -Overall competence (mCEX) vs. structured oral (RCSPC-IM) = .73 -Overall competence (mCEX) vs. bedside station sub score (RCSPC-IM) = .67 -Overall competence (mCEX) vs. written examination (RCSPC-IM) = .72 (2) Strong correlations between mCEX scores and corresponding ABIM MEF scores and ITE scores (3) For three clinical skills faculty participants were able to discriminate among performance level. Differences were statistically significant, however, the range in ratings among the participants for each videotape was wide. (4) The mini-CEX discriminates between pro-existing levels of global competency between residents: first year residents 7.19, second 7.51, third 7.76, fourth 8.16. This differences are statistical significant (P=.0008). (5) Values did not vary significantly between groups or testing periods, nor was the interaction significant. (6) Participants felt that the assessment methods were fair. With each method, more senior trainees received significant higher scores. (7) Nine-point scales appear to provide more accurate scores.
Mini-CEX in clerkship	Yes	Concurrent validity and criterion validity *(Kogan et al., 2003).*	Relationships with other assessments and an increase in scores with increasing clinical experience.	Significant correlation between mean of Mini-CEX scores and exam scores/ write ups/ inpatient, outpatient and final course grades. Students who received 'honours' on their exam had higher mean mini-CEX scores than those who received 'pass'. Mean mini-CEX ratings increased over the course of the year.
OCEX	Yes	(1) Construct validity *(Golnik & Goldenhar, 2005)* (2) Content validity *(Golnik et al., 2004).* (3) Face validity *(Golnik et al., 2004).*	(1) Not mentioned. (2&3) Modifying the OCEX in ways suggested by a panel of experts.	(1) They concluded good validity: In the subscale 'examination' and 'interview skills' they perhaps should not necessarily hang together. It might not be indicating a poorly functioning scale per se. The OCEX shows validity. (2&3) OCEX has face and content validity. It can be used to assess resident's patient care skills, medical knowledge, and interpersonal skills.

Appendix 5 Validity

	Validity determined?	Type(s) of validity	How was this performed?	Outcomes
Palliative care CEX	No	-	-	-
P-MEX	Yes	(1) Content validity *(Cruess et al., 2006).* (2) Construct validity *(Cruess et al., 2006).*	(1) Process of item generation, and processes compared to other groups in North America. (2) Factor analysis.	(1) Extensive process of item generation, and the results of this process were 'triangulated' with similar processes conducted by other groups in North America. Researchers conclude that this study suggests content validity. (2) The 24 original items cluster into identifiable factors or facets of the construct. This is evidence for construct validity.
CBA rheumatology	Yes	Face validity *(Dowson & Hassell, 2006).*	Not mentioned	Researchers conclude that it is clear that it will test the skills used in routine outpatient consultations, i.e. it has face validity. Researchers mentioned that if looked at in isolation, there are weaknesses in terms of the validity. The assessment described should be viewed as one of a selection of assessments utilized.
SCO	No	-	-	-
PEAF	No	-	-	-
GRFA	No	-	-	-
WRF	Yes	(1) Content validity *(Turnbull et al., 2000).* (2) Concurrent validity *(Turnbull et al., 2000).*	(1) Process of internal and external expert reviews. (2) Examining the correlations of scores on the different measures.	(1) The form did reflect the necessary domains of practice (is considered to have content validity). The high item-total correlation suggests that evaluators may not be reflecting these competencies. (2) Correlations between WRF and Admission Rating Form is moderate (.47). Correlation between WRF and Multidisciplinary Team Rating Forms were lower (.37 and .26 (history and physical forms)). Modest correlation between different forms provides further supportive evidence of validity.
CPB	No	-	-	-
CEE	No	-	-	-
CSAF	Yes	Construct validity *(Links et al., 1984).*	Improvement in skills from the pre- to post- direct observation exercise was measured. Different real patients, (in most cases) same supervisor. Paired t-test (sig. < .0025).	9 of 17 significant on .0025 level. 13 of 17 significant on .05 level.
SDOT	No	-	-	-
ECS	No	-	-	-

Appendix 5 Continued

	Validity determined?	Type(s) of validity	How was this performed?	Outcomes
LEP	Yes	Construct validity *(Prescott-Clements et al., 2008)*.	LEP results of 2 cohorts and rater questionnaires.	-Appropriate range of assessment. -Increase in ratings throughout the year.
MiniCard	Yes	Accuracy *(Donato et al., 2008)*.	Comparison of script	The new format is more accurate than the mini-CEX form.
CEC	Yes	Criterion validity *(Richards et al., 2007)*.	Pearson correlation coefficient	Significant positive correlation on CEC with 'clinical performance ratings' 'NBME' and 'final grades'. No positive correlation on CEC with 'OSCE scores'.
BFA	No	-	-	-

Appendix 5 Continued

	Effects on trainee's learning determined?	How was this performed?	Outcomes
Mini-CEX	Yes	(1) Questionnaires *(Cook et al., 2008)*. (2) Focus groups *(Malhorta et al., 2008)*.	(1) Although participants' confidence in using the mini-CEX improved, beliefs about the accuracy and usefulness of the mini-CEX changed little. (2) Repeated mini-CEX experiences made the resident learner more confident in controlling and enforcing the positive educational effect.
Mini-CEX in clerkship	No	-	-
OCEX	No	-	-
Palliative care CEX	Yes	Questionnaires (telephone survey and written questionnaire) *(Han et al., 2005)*.	After the intervention there is observed significant improvement in self-rated competence and confidence in giving bad news ($p < .05$). Researchers concluded that the study provides support for the educational value of the form.
P-MEX	Yes	Semi-structured interviews *(Cruess et al., 2006)*.	Useful in promoting self-reflection. Awareness of the importance of professionalism in daily encounters. Identifying behaviours consistent with professionalism. Teaching about these subjects matters.
CBA rheumatology	No	-	-
SCO	Yes	Questionnaires with faculty members *(Lane &Gottlieb, 2000)*.	Value of SCO to student education: 4 out of 5 on a 5-point scale. Researchers conclude that they have evidence of students incorporating feedback suggestions into future encounters either by actually observing them or by the student telling that they had done so. A quantitative evaluation on the educational impact of the program needs to be undertaken.

Appendix 6 Effect on trainee's learning

	Effects on trainee's learning determined?	How was this performed?	Outcomes
PEAF	No	-	-
GRFA	No	-	-
WRF	No	-	-
CPB	No	-	-
CEE	No	-	-
CSAF	Yes	Students (N=19) completed a ranking of the components of the clerkship. The components were ranked in terms of their contribution to the development of knowledge, skills and attitude *(Links et al., 1984)*. A comparison was made of the quality of written evaluation forms completed prior to introduction (N=23) and after its introduction (N=35). Rated on a 5-pointscale *(Links et al., 1984)*.	Students rated the direct observation exercise as the second most valuable component in terms of assisting skill acquisition. Quality rating before introduction 2.8 (range 2 to 4). Quality rating after introduction 3.6 (range 2 to 5) (sig .05).
SDOT	Yes	Opinion of SDOT task force *(Shayne et al., 2006)*.	In the opinion of the SDOT task force, specific observation of resident's behaviour and actions provide more substantive feedback for residents to understand and modify their performance.
ECS	No	-	-
LEP	Yes	Rater questionnaire *(Prescott-Clements et al., 2008)*.	In first cohort 79% of raters found positive response from trainees on feedback, in second cohort 86%
MiniCard	No	-	-
CEC	Yes	End of clerkship evaluations of intervention-group compared with historical control-group *(Paukert et al., 2002)*.	Improved satisfaction with feedback and improved understanding of history taking, physical examination and decision making skills.
BFA	Yes	Student questionnaires *(Burch et al., 2006)*.	95,6% recognised the learning value of BFA; 70% acknowledged the informative, advisory and motivational role of feedback; 71,9% increased in preparatory reading.

Appendix 6 Effect on trainee's learning

References

1. Wass, V., Vleuten, van der, C., Shatzer, J. & Jones, R. (2001). "Assessment of clinical competence". *Lancet, 357,* 945 – 949.

2. Norcini, J.J., Blank, L.L., Arnold G.K. & Kimball, H.R. (1995). "The mini-CEX (clinical evaluation exercise): a preliminary investigation". *Annals of Internal Medicine, 123,* 795 - 799.

3. Norcini, J.J., Blank, L.L., Duffy, F.D. & Fortna, G.S. (2003). "The mini-CEX: a method for assessing clinical skills". *Annals of Internal Medicine, 138,* 476 - 481.

4. Prescott, L.E., Norcini, J.J., McKinlay, P. & Rennie, J.S. (2002). "Facing the challenges of competency-based assessment of postgraduate dental training: Longitudinal Evaluation of Performance (LEP)". *Medical Education, 36,* 92 - 97.

5. Lane, J.L. & Gottlieb, R.P. (2000). "Structured clinical observations: a method to teach clinical skills with limited time and financial resources". *Pediatrics, 105,*973 - 977.

6. Paukert, J.L., Richards, M.L. & Olney, C. (2002). "An encounter card system for increasing feedback to students". *The American Journal of Surgery, 183,* 300 - 304.

7. Vleuten, van der, C.P.M. (1996). "The assessment of professional competence: developments, research and practical implications". *Advances in Health Sciences Education, 1,* 41 – 67.

8. Kogan, J.R., Holmboe, E.S. & Hauer, K.E. (2009). "Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review", *JAMA, 302,* 1316-1326.

9. Wass, V. & Vleuten, van der, C. (2004). "The long case". *Medical Education, 38,* 1176 - 1180.

10. Alves de Lima, A., Barrero, C., Baratta, S., Castillo Costa, Y., Bortman, G., Carabajales, J., Conde, D., Galli, A., Degrange, G. & Vleuten, van der, C. (2007). "Validity, reliability, feasibility and satisfaction of the mini-clinical evaluation exercise (mini-CEX) for cardiology residency training". *Medical Teacher, 29,*785 - 790.

11. Anderson, C.I., Jentz, A.B., Harkema, J.M., Kareti, L.R., Apelgren, K.N. & Slomski, C.A. (2005). "Assessing the competencies in general surgery residency training". *The American Journal of Surgery,189,*288 - 292.

12. Burch, V.C., Seggie, J.L. & Gary, N.E. (2006). "Formative assessment promotes learning in undergraduate clinical clerkships". *South African Medical Journal, 96,*430 - 433.

13. Cook, D.A. & Beckman, T.J. (2009). "Does scale length matter? A comparison of nine- versus five-point rating scales for the mini-CEX". *Advances in Health Science Education, 14,* 655-664.

14. Cook, D.A., Dupras, D.M., Beckman, T.J., Thomas, K.G. & Pankratz, S. (2008). "Effect of rater training on reliability and accuracy of mini-CEX scores: a randomized, controlled trial". *Journal of General Intern Medicine, 24,* 74 - 79.

15. Cruess, R., McIlroy, J.H., Cruess, S., Ginsburg, S. & Steinert, Y. (2006). "The professionalism mini-evaluation exercise: a preliminary investigation". *Academic Medicine, 81,* S74 - S78.

16. Donato, A.A., Pangaro, L., Smith, C., Rencic, J., Diaz, Y., Mensinger, J. & Holmboe, E. (2008). "Evaluation of a novel assessment form for observing medical residents: a randomised, controlled trail". *Medical Education,42,* 1234 - 1242.

17. Dowson, C. & Hassell, A. (2006). "Competence-based assessment of specialist registrars: evaluation of a new assessment of out-patient consultations". *Rheumatology, 45,* 459 - 464.

18. Durning, S.J., Cation, L.J., Markert, R.J. & Pangaro, L.N. (2002). "Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training". *Academic Medicine,77,* 900 - 904.

19. Golnik, K.C., Goldenhar, L.M., Gittinger, J.W. & Lustbader, J.M. (2004). "The ophthalmic clinical evaluation exercise (OCEX)". *Opthalmology, 111,* 1271 - 1274.

20. Golnik, K.C. & Goldenhar, L. (2005). "The ophthalmic clinical evaluation exercise: reliability determination". *Opthalmology,112,* 1649 - 1654.

21. Han, P.K., Keranen, L.B., Lescisin, D.A. & Arnold, R.M. (2005). "The palliative care clinical evaluation exercise (CEX): an experience-based intervention for teaching end-of-life communication skills". *Academic Medicine, 80,* 669 - 676.

22. Hatala, R., Ainslie, M., Kassen, B.O., Mackie, I. & Roberts, J.M. (2006). "Assessing the mini-clinical evaluation exercise in comparison to a national specialty examination". *Medical Education, 40,* 950 - 956.

23. Hatala, R. & Norman, G.R. (1999). "In-training evaluation during an internal medicine clerkship". *Academic Medicine, 74,* S118 - S120.

24. Holmboe, E.S., Huot, S., Chung, J., Norcini, J. & Hawkins, R.E. (2003). "Construct validity of the mini clinical evaluation exercise (miniCEX)". *Academic Medicine, 78,* 826 - 830.

25. Kogan, J.R., Bellini, L.M. & Shea, J.A. (2002). "Implementation of the mini-CEX to evaluate medical students' clinical skills". *Academic Medicine, 77,* 1156 - 1157.

26. Kogan, J.R., Bellini, L.M. & Shea, J.A. (2003). "Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship". *Academic Medicine, 78,* S33 - S35.

27. Kogan, J.R. & Hauer, K.E. (2006). "Brief report: Use of the mini-clinical evaluation exercise in internal medicine core clerkships". *Journal of General Internal Medicine, 21,* 501 - 502.

28. Links, P.S., Colton, T. & Norman, G.R. (1984). "Evaluating a direct observation exercise in a psychiatric clerkship". *Medical Education, 18,* 46 - 51.

29. Malhotra, S., Hatala, R. & Courneya, C.A. (2008). "Internal medicine residents' perceptions of the mini-clinical evaluation exercise". *Medical Teacher, 30,* 414 - 419.

30. Margolis, M.J., Clauser, B.E., Cuddy, M.M., Ciccone, A., Mee, J., Harik, P. & Hawkins, R.E. (2006). "Use of the mini-clinical evaluation exercise to rate examinee performance on a multiple-station clinical skills examination: a validity study". *Academic Medicine, 81,* S56 - S60.

31. Nair, B.R., Alexander, H.G., McGrath, B.P., Parvathy, M.S., Kilsby, E.C., Wenzel, J., Frank, I.B., Pachev, G.S. (2008). "The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates". *The Medical Journal of Australia, 189,* 159 - 161.

32. Norcini, J.J., Blank, L.L., Arnold, G.K. & Kimball, H.R. (1997). "Examiner differences in the mini-CEX". *Advances in Health Science Education, 2,* 27 - 33.

33. Norcini, J. & Burch, V. (2007). "Workplace-based assessment as an educational tool: AMEE Guide No. 31". *Medical Teacher, 29,* 855 - 871.

34. Nyman, K.C. & Sheridan, B. (1997). "Evaluation of consulting skills of trainee general practitioners". *Australian Family Physician, 26,* S28 - S35.

35. Prescott-Clements, L., Vleuten, van der, C.P.M., Schuwirth, L.W.T., Hurst, Y. & Rennie, J.S. (2008). "Evidence for validity within workplace assessment; the Longitudinal Evaluation of Performance (LEP)". *Medical Education, 42,* 488 - 495.

36. Richards, M.L., Paukert, J.L., Downing, S.M. & Bordage, G. (2007). "Reliability and usefulness of clinical encounter cards for a third-year surgical clerkship". *Journal of Surgical Research, 140,* 139 - 148.

37. Ringsted, C., Ostergaard, D., Ravn, L., Pedersen, J.A., Berlac, P.A. & Vleuten, van der, C.P.M. (2003). "A feasibility study comparing checklists and global rating forms to assess resident performance in clinical skills". *Medical Teacher, 25,* 654 - 658.

38. Ross, R. (2002). "A clinical-performance biopsy instrument". *Academic Medicine, 77,* 268.

39. Shayne, P., Gallahue, F., Rinnert, S., Anderson, C.L., Hern, G. & Katz E. (2006). "Reliability of a core competency checklist assessment in the emergency department: the standardized direct observation assessment tool". *Academic Emergency Medicine, 13,* 727 - 732.

40. Shayne, P., Heilpern, K., Ander, D. & Palmer-Smith, V. (2002). "Protected clinical teaching time and a bedside clinical evaluation instrument in an emergency medicine training program". *Academic Emergency Medicine, 9,* 1342 - 1349.

41. Torre, D.M., Simpson, D.E., Elnicki, D.M., Sebastian, J.L. & Holmboe, E.S. (2007). "Feasibility, reliability and user satisfaction with a PDA-based mini-CEX to evaluate the clinical skills of third-year medical students". *Teaching and Learning in Medicine, 19,* 271 - 277.

42. Turnbull, J., MacFadyen, J., Barneveld, van, C. & Norman, G. (2000). "Clinical work sampling. A new approach to the problem of in-training evaluation". *Journal of General Internal Medicine, 15,* 556 - 561.

43. Wilkinson, J.R., Crossley, J.G.M., Wragg, A., Mills, P., Cowan, G. & Wade, W. (2008). "Implementing workplace-based assessment across the medical specialties in the United Kingdom". *Medical Education, 42,* 364 - 373.

44. Vleuten, van der, C. P.M. & Schuwirth, L.W.T. (2005). Assessing professional competence: from methods to programmes. *Medical Education, 39,* 309-317.

Chapter 3

The process of feedback in workplace-based assessment: organization, delivery, continuity

Abstract

Objectives Feedback in workplace-based clinical settings often relies on expert trainers' judgements of directly observed trainee performance. There is ample literature on effective feedback, but in practice trainees in workplace-based training are not regularly observed. We aimed to examine external conditions that impact feedback in observational workplace-based assessment (WBA).

Methods Interviews were conducted and resulting data analysed using a qualitative, phenomenological approach. Between October 2009 and January 2010, we interviewed 22 postgraduate general practice trainees at two institutions in the Netherlands. Three researchers analysed the transcripts of the interviews.

Results A three-step scheme emerged from the data. Feedback as part of WBA is of greater benefit to trainees if: (1) observation and feedback are planned by trainee and trainer; (2) the content and delivery of the feedback are adequate; (3) the trainee uses the feedback to guide his or her learning by linking it to learning goals. Negative emotions reported by almost all trainees in relation to observation and feedback led to different responses. Some trainees avoided observation, whereas others overcame their apprehension and actively sought observation and feedback. Active trainers where able to help trainees overcome their fears. Four types of trainer-trainee pairs were distinguished according to their engagement in observation and feedback. External requirements set by training institutions may stimulate inactive trainers and trainees.

Conclusions In line with the literature our results emphasise the importance of the content of feedback and the way it is provided, as well as the importance of its incorporation in trainees' learning. Moreover, we highlight the step before the actual feedback itself. The way arrangements for feedback are made appears to be important to feedback in formative WBA. Finally, we outline several factors that influence the success or failure of feedback but precede the process of observation and feedback.

Introduction

Feedback on trainees' performance in workplace-based clinical settings generally reflects the judgement of experts who have observed the performance.[1] The intended effect of this feedback is to help trainees learn and improve their performance and is based on the assumption that feedback creates awareness of shortcomings and thereby motivates learners to improve or change.[2]

This study focuses on formal feedback that is based on the observation of a trainee conducting a consultation with a single patient. Although frequent observations also provide information about performance for the overall assessment of trainees, we focus in this study on the formative aspect of feedback, whereby trainees purposefully invite and receive feedback in order to learn and improve from it. Many articles have reported on when and how feedback is likely to be most effective,[1-4] but there are strong concerns that trainees only infrequently receive feedback based on direct observation of a patient encounter.[5-7] We will first present an overview of the literature that is relevant to the scope of the study and will then examine practice in the clinical setting.

Most of the studies on feedback on performance have focused upon factors that influence the use of feedback and strategies for delivering it.[3,4] In a review, Shute showed that the actual use of feedback in different settings depends on motivation (the trainee needs the feedback), opportunity (feedback is given in time to be used by the trainee) and means (the trainee is able and willing to use the feedback).[3] Shute also provides comprehensive lists of guidelines to enhance learning.[3] Some of the implications of the guidelines are that feedback should be given in response to a problem or task, may prevent or correct errors and misconceptions, should preferably be written and should promote a specific learning goal. It should also be objective, which means that effective feedback will allow for a comparison of performance with an established standard of performance. In a review of the literature on the impact of feedback on learning and achievement, Hattie and Timperley[4] reported that effective feedback should focus on the task, process and (self-)regulation. Feedback relating to the personal level is rarely effective because trainees are very strongly focused on avoiding risk and failure. In their model of feedback, Hattie and Timperley also emphasised goal orientation exemplified by three questions that a trainee should be able to answer if feedback is to be effective: Where am I going? How am I going? Where do I go to next?[4] The model gives a good overview of how feedback might enhance learning. Both reviews also showed that the type and delivery of feedback influence its effectiveness and that it may even have a negative impact.[3,4]

A review of feedback within the medical domain, by Archer[8], indicated that for feedback to be relevant and effective, it should be specific and not exclusively trainer driven. It should be part of a two-way process and trainees should be given the

opportunity to reflect on their actions. In Archer's model of effective feedback, feedback is not a series of unrelated events, but a sequential process linked to personal goals that includes self-monitoring (reflection on action) that is supported by external feedback[8].

In a study of multi-source feedback for practising doctors, Sargeant et al.[9] found that acceptance and usage of feedback by individual doctors depended on the nature of the feedback. Positive feedback appeared to be easily assimilated by recipients, whereas negative feedback was first appraised for its credibility based on its process (feedback based on observed performance), source, content and specificity, as well as on its congruence with feedback from other sources. Watling and Lingard also concluded that participants' perceptions of an evaluation process profoundly affected the usefulness of the evaluation and the extent to which it achieved its goal[1]. Similarly, Eva et al.[10] found that the self-perceptions of recipients of feedback, relating to their confidence, experience and fear of not being sufficiently knowledgeable, impacted on their interpretation and uptake of feedback.

According to Norcini and Burch,[5] there are indications that trainees are observed only rarely and that faculty staff play a critical role in the successful implementation of formative assessment.[5,6] By contrast, in their review, Miller and Archer[2] pointed to evidence that formative feedback in the workplace is highly appreciated by users, who (subjectively) reported positive education impact. How is this positive subjective response of users reconciled with Norcini and Burch's claim[5] that observation and feedback happen only too rarely?

In summary, with respect to formal feedback based on the observation of a trainee during a consultation with a single patient, we known that formative feedback in workplace-based clinical settings relies on the judgement of expert trainers and that the literature has provided models, lists and guidelines indicating when feedback is likely to be effective and how it is best provided. However, there seems to be a discrepancy between the evidence in the literature and what we see in day-to-day practice, in which the application of the evidence with respect to feedback falls short of what we would expect. To clarify this issue, we conducted a qualitative study in which we explored the lived experiences and perceptions of individual trainees by addressing the following research questions:

1. How is feedback, based on observation of a trainee performing a consultation with a single patient, carried out in the workplace-based clinical setting?
2. Which aspects influence the feedback process in the workplace-based clinical setting?

Methods

Because we wanted to be receptive to all experiences and perspectives, we conducted semi-structured individual interviews to collect data. We conducted the study in the clinical setting of postgraduate training for general practice.

Context
From the eight departments of general practice in the Netherlands that offer postgraduate training, we selected two for inclusion in the study. General practice trainees learn by working in one general practice for a prolonged period under the supervision of a general practitioner (GP) who works in the practice. Trainees work mostly independently, but can ask their trainer for help and advice. There are no external rules to guide observation of trainee performance. Both the trainer and trainee can make arrangements for the observation of a consultation with feedback by planning a consultation at which they are both present or by recording a consultation and planning a meeting for observation and feedback sometime afterwards.

The three-year postgraduate training programme for general practice consists of two years in general practice (years 1 and 3) and one year (year 2) of rotations in hospitals and other medical institutions. The eight training institutions deliver a national programme in ways that are broadly similar but leave room for local interpretation. National summative examinations include knowledge tests and work-based communication video-based assessment. Formative assessments are organised locally. We conducted the study in two departments of general practice which differ in their organisation of and recommendations to trainees in relation to (formative) feedback based on observation in daily practice. We selected two departments to substantiate our findings. The Nijmegen programme recommends the direct observation of performance combined with a mini-clinical exercise (mini-CEX) instrument that covers the competencies of medical expertise, communication and professionalism. The instrument allows for written narrative reflections and feedback based on the Pendleton rules,[11] with reflection on 'what went well', feedback on 'what went well', reflection on 'what could have been done better', feedback on 'what could have been done better' and the planning of further activities. Trainers and trainees are advised to conduct one feedback session using the mini-CEX instrument per week and to conduct a minimum of three sessions every three months. The results can be used in trainees' three-monthly progress interviews and thus the feedback is not strictly formative because when it is used frequently, it gives the trainer input for summative purposes. Compliance with the recommendations is not monitored. In Maastricht, videotaped observation combined with a feedback discussion is recommended. The daily learning meetings of trainer and trainee can be used to watch and discuss a videotape of a patient encounter. This

discussion does not have to be supported by a tool, but the format of the national communication video-based assessment can be used if desired. No minimum has been set for the number of observations. The recommendations of the institutions are directive, but trainers and trainees are free to organise observation differently. Although the institutions' recommendations for observation differ, the feedback sessions are of comparable intensity and both generate substantial narrative feedback. Both departments see the primary aim of feedback as supporting the learning of trainees, although it can also be used as input for the portfolios and to inform trainers about a trainee's competence for the purposes of progress decisions.

Participants

Because we wanted to explore trainees' lived experiences with regard to observation and feedback on performance in general practice, we interviewed eleven first-year and eleven third-year general practice trainees in the period between October 2009 and January 2010. E-mails were sent to general practice trainees in year 1 and 3 at both Nijmegen and Maastricht to invite them to participate. Participation was voluntary. Of the 27 trainees who responded, we selected a purposeful sample of 22 trainees, based on institution, gender and year of education. Because of this sampling and for practical reasons, we scheduled and conducted all 22 interviews. Participants received a gift coupon to the value of €20.00 and were assured that all data would be used anonymously and confidentially. Participants gave informed consent, which was confirmed on the audiotape during the interview. The study was approved by the ethics review board of the Dutch Association for Medical Education (Nederlandse Vereniging voor Medisch Onderwijs [NVMO]).

Design

The interviews lasted 20-30 minutes and were all conducted by the first author (EP). Based on the literature and the results of two pilot interviews (conducted in July and September 2009), we designed an interview guide that covered the topics of: practical organisation; frequency; receiving and accepting feedback, and the relationship of observations to other assessment formats included in the portfolio. The interview guide is shown in Appendix 1.

Data analysis

Because of the exploratory nature of the study and our focus on trainees' experiences, we used a qualitative approach related to phenomenology, which is not an empirical analytical science, but a philosophy looking for the meaning of phenomena.[12-13] Our study draws on the assumptions of what has come to be known as 'new' or 'American' phenomenology, which can also include interpretation of data. The aim is to describe participants' 'lived experiences within the context' in order to find a general meaning.[13]

We audio-recorded and literally transcribed the interviews in a manner that safeguarded the anonymity of participants and their trainers. Firstly, all interviews were read by EP to gain an overall impression of the content and possible themes. Then, one interview was read and themed by three researchers (EP, HM, AK). They discussed the emergent themes in order to ascertain that they were drawing the same concepts from the transcripts and to establish how they might formulate codes. An example of a theme is: 'routine in observation'. Possible codes associated with this theme are: 'Wednesday morning' and 'Thursday afternoon'. Five transcripts were coded by EP and HM, and five other transcripts were coded by EP and AK. This resulted in codes and themes for eleven interview transcripts. Atlas-ti was used to manage the codes and themes. The discussion about the themes and codes from the first eleven interviews and the analysis of the remaining eleven interviews were considered to provide validation of the analysis. The second set of eleven transcripts was coded mainly by EP, but, to ensure consistency in the coding, three interviews in this set were coded by EP, HM and AK, EP and HM and EP and AK respectively. Saturation was reached after 13 interviews had been coded (no new themes emerged), but all 22 interviews were analysed and the remaining nine interviews were used to confirm saturation. Data from all 22 interviews were consistent with the results.

The interviews were conducted and transcribed in Dutch. The quotations in the results section are literal translations of the original statements.

Results

Three steps appeared to be necessary for feedback to have a beneficial effect during single-encounter assessments. The first step concerned arrangements for observation and feedback made by trainer and trainee together. The second step related to the content and delivery of the feedback. The third step concerned the incorporation of the feedback in the learning process and required the trainee to accept the feedback, reflect on it in relation to his or her learning goals, and use it to plan some kind of action to pursue these learning goals.

We combined these three steps into a scheme (Figure 1). We will discuss the three steps consecutively and then answer the second research question by showing which aspects influenced the extent to which the steps of the scheme were actually used. Finally, we will discuss the effects of the different approaches recommended by the two institutions.

Step 1: Organisation of observation and feedback

The frequency of observation of consultations varied considerably and ranged from twice a weekly to not at all. Remarkably, trainees who reported a high observation

Figure 1 Steps in the process of obtaining useful feedback

rate also reported having made agreements with their trainers on clearly defined training routines at the beginning of the training year:

We have an arrangement that we just, say during consultation hours, that she observes me every Thursday: my last two consultations on Thursday afternoons (Trainee 16).

Trainers and trainees who had set clear routines for observation also scheduled time for feedback immediately following direct observation or in the same week in the case of videotape observation. These routines appeared to promote the effective completion of the first step.

Trainees spontaneously mentioned observation during home visits, and weekend duty or night shifts. These occasions created natural moments for observation because trainers and trainees were generally working together. Trainees remarked, however, that the content of these encounters and, consequently, of the feedback differed from that of 'regular practice'. This meant that these natural moments for observation could not replace observation and feedback on 'regular consultations'.

Step 2: Content and delivery of feedback
The analysis showed that most of the feedback related to communication. Even with the use of the mini-CEX instrument, which specifically invited feedback on three competencies (medical expertise, communication and professionalism), communication was the predominant topic of feedback.

There was considerable variety in the delivery of feedback. At one end of the spectrum were very intensive sessions in which specific feedback was given and trainees reflected on their performance and considered further action to improve it. One trainee described this as:

We frequently stop it (the video). And then we watch... mostly he first asks me what I think of it, and what I think I could have done differently, or should have done differently (...). And the effect is that we frequently role-play in between (Trainee 4).

At the other end of the spectrum lay superficial, non-specific feedback in which no attention was paid to reflection or further action. One trainee reported:

My GP trainer (…) said twice: "Yes (…) I have nothing to do right now"; two or three times. "I can sit in with you". And at the end of the consultation: "Rather fine, I would not have done it differently, seems accurate". (Trainee 14)

Step 3: Incorporation of feedback into the learning process
Some trainees proceeded from the first and second step to incorporation of the feedback into their learning. These trainees were aware of the relationship of feedback with their learning goals, their portfolio and the results of other assessments.

We used it because in my portfolio, due to, yeah to find out how you explore the 'request for help', and we did that with a mini-CEX (Trainee 20).

Trainees used observation and feedback to judge their progress with regard to learning goals formulated earlier, and they also used feedback to formulate new learning goals. When a trainee had completed all three steps, the cycle was able to start again when observation was planned to elicit further feedback on the same learning goals or on newly formulated goals. This process is represented by the large arrow pointing from step three back to step one in Figure 1.

Factors influencing the process of receiving feedback
Not all trainees used all three steps: some failed to take the first step or to proceed to the second or third steps. To answer our second research question, we investigated the factors influencing this process.

Attitudes towards observation and feedback
The analysis of the interviews showed that trainees' reactions to observation were primarily emotional. They talked spontaneously about feeling apprehensive about observation, saying that they felt they behaved differently when they were being observed. They also said they did not like being observed. Both videotaped and direct observation were considered to interfere with normal practice.

It is intuitive, if someone watches you, I get nervous. It is not that my GP-trainer is doing, or not doing, something, but generally... . (Trainee 10)

Although almost all trainees mentioned these emotions, two distinct patterns of responding to them emerged. Some trainees wanted to be observed and actively sought feedback, despite their fears. They acknowledged that they could benefit

from feedback and that it could help them to improve their performance. Other trainees, however, allowed their fears to prevail and were reluctant to make arrangements for observation and feedback. As observation and feedback were not subject to external control (other than being strongly recommended), these trainees were able to avoid them.

Avoidance behaviour could be counteracted by an active GP-trainer. Based on the data from the interviews, two groups of trainers were identified: one set of trainers took initiatives to ensure that observation and feedback took place, whereas the other set refrained from observation and feedback.

My trainer said: "(…) shouldn't we do something like a mini-CEX?" I said: "yes". And then he said: "With my last trainee we used the walk-in surgery (…)" (Trainee 1)

I think it (observation) is useful, but also scary, I think that's the reason I am not inclined to mention it to my trainer. But if you just do it, then it gets easier (…) and then it is very helpful. (Trainee 1)

I think it is just laziness that I don't do it. And I notice that my GP-trainer does not ask for it either. And of course that is a little childish (of me). (Trainee 5)

Most trainers who actively observed their trainees also invited their trainees to observe them in order to provide a learning experience for the trainee. Based on the attitudes of trainees and trainers towards observation, four groups of trainee-trainer pairs were distinguished (Table 1).

Recommendations from the training institution

In addition to trainee and trainer attitudes, the recommendations of the training institution influenced the occurrence of observation and feedback. The institutional recommendations (direct observation using the mini-CEX format at least three times in three months [Nijmegen]; videotaped observation with a feedback session [Maastricht]) influenced the feedback process in various ways. Some trainees reported that they were observed because the institution required it, whereas other trainees said they would organise more observations if a minimum number was required and closely monitored by the institution. By contrast, trainer-trainee pairs who showed a positive attitude towards observation and feedback regarded the imposition of strict requirements as excessive regulation:

I think it depends on the relationship with the trainer. If things are going well, he (the trainer) knows how you work and how you are doing. But if things aren't going well, then you can use it and say: hey, we need to do this (mini-CEX). (Trainee 15)

	GP-trainer shows active behaviour in relation to giving feedback	GP-trainer does not show active behaviour in relation to giving feedback
Trainee shows active feedback-seeking behaviour	Frequent observation and feedback.	GP-trainer does not take the initiative to give feedback. When the trainee asks for feedback it may or may not be provided.
Trainee does not show active feedback-seeking behaviour	Trainee does not seek observation and feedback, but they take place because the trainer takes action.	Observation and feedback only if externally required (by Training institution)

Table 1 Types of trainer-trainee pairs based on responses to negative emotions relating to feedback

Differences among trainees

We interviewed trainees from two different institutions, each of which recommended a different approach to formative feedback on observed consultations. Nevertheless, in practice, trainees in both institutions used both direct and videotaped observation. The mini-CEX instrument was used only in Nijmegen, where videotaped observation was also used. In Maastricht, video-based assessment is standard, but some trainees were also observed directly. The main reason for selecting one of the observation methods referred to the preferences of the trainee or trainer and the method's practical feasibility in general practice. We found no striking differences between Nijmegen and Maastricht. Trainees differed in their appreciation of the two methods, but selected the method they preferred. Appreciation depended on the degree of apprehension of the trainee and the personality of the GP-trainer. No differences emerged between trainees of different genders or years of education. Trainees in each category (institution, gender, year of education) showed no differences in whether they completed the three steps or failed to take the first, second or third step.

Discussion

Although the literature on feedback has mainly focused on the content and delivery of feedback[3,4,9] and on the recommendation that feedback should promote a learning goal [3,4,8] (the second and third steps of our scheme), our results underscore the importance of the step before the actual provision of feedback in the workplace setting. Deliberate planning of observation and feedback appears to be essential. The feedback literature explains how feedback can be effective, based on the assumption

that feedback takes place, but Norcini and Burch[5] and the present results show that this assumption may be rather unrealistic.

Although the second step in our scheme has been described earlier, our findings confirm that it is an important aspect of the feedback process and that its occurrence should not be taken for granted; some trainees take the first step of the scheme, but receive no effective feedback at all or not in an appropriate manner. With respect to the third step, Archer[8] stated that feedback should not comprise a series of unrelated events, but should be incorporated into the overall learning process by relating it to learning goals and plans for improvement. In line with this, Hattie and Timperley[4] highlighted the need to resolve the question of 'where to next' in order to make feedback effective. In addition, Shute[3] showed that formative feedback should promote an orientation based on learning goals. We think that feedback based on the observation of a consultation in formative workplace-based assessment (WBA) should not be an isolated event, but should represent the starting point of a continuing learning process. It can also be used to reflect on an ongoing learning goal.

Although feedback is a concept with many facets, the three-step scheme we propose can be useful in pinpointing exactly where things go wrong for trainees and thereby may prevent failure. Our second research question indicated our wish to further examine factors influencing the successful completion of the steps.

Apprehension about being observed and receiving feedback proved to have a powerful negative effect on feedback. The role of fear has also been described in the literature[14]. We found, however, that apprehension can be overcome when trainees are motivated to actively seek feedback. Teunissen et al[15] found a similar pattern, showing that trainees are not merely passive recipients of feedback, but differ in their ways of seeking or avoiding it. Our results showed that not only trainee motivation, but also the role of the trainer can help to overcome avoidance patterns induced by negative emotions. Future research should further examine this area, particularly in light of the substantial added value of trainer-led initiatives shown in this study. A trainer who actively promotes observation and feedback may be able to counteract trainees' avoidance behaviour with regard to feedback.

External regulation by the institutions also influenced the feedback process. The literature shows, however, that simply introducing a tool does not suffice to ensure good feedback that tie in with personal goals[16,17].

Our findings show that it remains important for trainers and trainees to actively pursue observation and feedback. Nevertheless, it may be advisable for institutions to enforce recommendations more strictly, given that our results indicate that it is quite easy for trainees and trainers not to comply with institutional recommendations at present. This suggests that the implementation of requirements for observation and feedback is a prerequisite for the provision of feedback in formative WBA. Possible ways to enhance implementation might include the provision of instructions and

training for trainees and trainers, stipulating a mandatory frequency of observation and feedback, and conducting a quality review.

The factors described here underscore the complex relationship between the feedback receiver, the feedback giver, the training institution and probably other environmental influences as well. Our explorative qualitative study has resulted in a framework of important elements that should be taken into account with regard to feedback on observed consultations in daily practice, but we do not claim that it offers comprehensive coverage of all possible influencing factors. More research is required to further investigate variables, the impact of variables and their interactions. In addition, we focused on the formative purpose of observation and feedback, but this is not as absolute as outlined. The process is not strictly formative because when it is used frequently, the outcomes can be used by trainers as input for summative purposes.

The prominence of communication as a topic of feedback may be attributed to the strong emphasis placed on communication skills in general practice training. Another explanation may be that summative performance assessment of communication skills during general practice training in the Netherlands is based on videotaped observations and several trainees use this method to obtain formative feedback.

There are some limitations to our study. The fact that participants were self-selected may have introduced bias. We do not know if participating trainees differed from non-participating trainees in their ways of seeking and organising observation and feedback. However, our results show wide variations in trainee experiences and ideas about observation and feedback. Furthermore, the transferability of the present findings to other medical specialties and work-based settings may be limited because the study was confined to general practice trainees.

An important finding, which can be interpreted as contributing to the theory on feedback, concerns the strong impact of the organisation and arrangements made for observations and feedback. Our results show that this crucial first step in the feedback process in formative WBA is influenced by the individual characteristics of both trainer and trainee, as well as by external regulations imposed by the training institution. We therefore recommend that institutions set a mandatory minimum number of observation and feedback sessions to be completed, supported by a mini-CEX instrument to direct feedback content and help trainees to connect the feedback with their learning goals. Moreover, institutions should train the trainers to make better use of observation as a teaching aid. Training could be structured around the three steps indentified in this study. We also recommend further research into trainer behaviour in the provision of feedback in order to investigate how trainers perceive their role in the feedback process and how trainees and trainers influence one another's feedback-giving and feedback-seeking behaviours.

References

1. Watling, C. J. & Lingard, L. (2012). Toward meaningful evaluation of medical trainees: the influence of participants' perceptions of the process. *Advances in Health Sciences Education, 17,* 183 – 194.
2. Miller, A. & Archer, J. Impact of workplace-based assessment on doctors' education and performance: a systematic review. *British Medical Journal, 341,* c5064, doi: 10.1136/bmj.c5064.
3. Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78,* 153 - 189.
4. Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77,* 81 - 112.
5. Norcini, J. & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher, 29,* 855 – 871.
6. Velden, van der, L.F.J & Batenburg, R.S. De opleiding tot huisarts opnieuw beoordeeld. Een onderzoek onder huisartsen in opleiding en alumni. http://www.nivel.nl/pdf/Rapport-kwaliteit-huisartsenoplei-ding.pdf. [Accessed 16 March 2012].
7. Burdick, W.P. & Schoffstall, J. (1995). Observation of emergency medicine residents at the bedside: how often does it happen? *Academic Emergency Medicine, 10,* 909 – 913.
8. Archer, J.C. (2010). State of the science in health professional education: effective feedback. *Medical Education, 44,* 101 – 108.
9. Sargeant, J., Mann, K., Sinclair, D., Vleuten, van der, C. & Metsemakers, J. (2007). Challenges in multisource feedback: intended and unintended outcomes. *Medical Education, 41,* 583 - 591.
10. Eva, K.W., Armson, H., Holmboe, E., Lockyer, J., Loney, E., Mann, K. & Sargeant, J. (2012) Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education, 17,* 15 - 26.
11. Pendleton, D., Schofield, T., Tate, P. & Havelock, P. (1984). *The Consultation: An Approach to Learning and Teaching.* Oxford: Oxford University Press, 68 – 71.
12. Kvale, S. (2007) *Doing interviews.* Los Angeles, CA: Sage Publications 10 – 21.
13. Dowling, M. (2007). From Husserl to van Manen. A review of different phenomenological approaches. *International Journal of Nursing Studies, 44,* 131 – 142.
14. Värlander, S. (2008). The role of students' emotions in formal feedback situations. *Teaching in Higher Education, 13,* 145 – 156.
15. Teunissen, P., Stapel, D., Vleuten, van der, C., Scherpbier, A., Boor, K. & Scheele, F. (2009). Who wants feedback? An investigation of the variables influencing residents' feedback-seeking behavior in relation to night shifts. *Academic Medicine, 84,* 910 – 917.
16. Holmboe, E.S., Yepes, M., Williams, F. & Huot, S.J. (2004). Feedback and the mini clinical evaluation exercise. *Journal of General Internal Medicine, 19,* 558 – 561.
17. Pelgrim, E.A.M., Kramer, A.W.M., Mokkink, H.G.A. & Vleuten, van der, C.P.M. (2012). Quality of written narrative feedback and reflection in a modified mini-clinical evaluation exercise: an observational study. *BMC Medical Education, 12,* 97

APPENDIX 1

Items included in the semi-structured interviews.

1. How many times does your trainer observe you (with or without a mini-CEX instrument)?

   Is this direct observation or video-based observation?

   In which situations does your trainer observe you?

   Who takes the initiative?

   Further exploration of the observation context for this specific trainee.

   And if not, why not? Exploration of the reasons for the absence of observation.

2. Do you receive feedback after these observations?

   When (do you receive feedback)?

   What is said?

   How does your trainer give feedback?

   Is there an opportunity for you to reflect on the feedback?

   What do you do with the feedback afterwards?

   Give a recent example of the feedback process in your training practice.

   Further exploration of the feedback context for this specific trainee.

   And if not, why not? Exploration of the reasons for the absence of feedback.

Chapter 4

Quality of written narrative feedback and reflection in a modified mini-clinical evaluation exercise: an observational study

Abstract

Background Research has shown that narrative feedback, (self) reflections and a plan to undertake and evaluate improvements are key factors for effective feedback on clinical performance. We investigated the quantity of narrative comments comprising feedback (by trainers), self-reflections (by trainees) and action plans (by trainer and trainee) entered on a mini-CEX form that was modified for use in general practice training and that encouraged trainers and trainees to provide narrative comments. In view of the importance of specificity as an indicator of feedback quality, we additionally examined the specificity of the comments.

Method We collected and analysed modified mini-CEX forms completed by GP trainers and trainees. Since each trainee has the same trainer for the duration of one year, we used trainer-trainee pairs as the unit of analysis. We determined for all forms the frequency of the different types of narrative comments and rated their specificity on a three-point scale: specific, moderately specific, not specific. Specificity was compared between trainee-trainer pairs.

Results We collected 485 completed modified mini-CEX forms from 54 trainees (mean of 8.8 forms per trainee; range 1-23; SD 5.6). Trainer feedback was more frequently provided than trainee self-reflections, and action plans were very rare. The comments were generally specific, but showed large differences between trainee-trainer pairs.

Conclusion The frequency of self-reflection and action plans varied, all comments were generally specific and there were substantial and consistent differences between trainee-trainer pairs in the specificity of comments. We therefore conclude that feedback is not so much determined by the instrument as by the users. Interventions to improve the educational effects of the feedback procedure should therefore focus more on the users than on the instrument.

Background

Research on formative assessment and feedback suggests that these are powerful tools to change trainees' behaviour[1-4]. Formative assessment is an instructional intervention evaluating performance and identifying trainees' strengths and weaknesses[1,5] in order to reveal performance gaps – i.e. differences between desired and actual performance[6]. From several studies we know that trainees do not benefit from feedback in the form of numerical marks[1,7,8], but that feedback should preferably be narrative and specific, explicating where more work needs to be done. Additionally, feedback can be made more effective when recipients receive guidance on how to turn feedback into concrete steps to improve their performance. Positive effects of narrative feedback have been reported by various authors, including Overeem et al.[9], who found higher satisfaction with such feedback, and Govaerts et al.[10], who suggested that narrative feedback can improve in-training evaluation. According to Sargeant et al.[11], feedback that is more specific is more readily assimilated, a view supported by Archer[12], who additionally concluded that feedback should not be exclusively trainer-driven but a two-way process in which trainers provide comments and at the same time encourage trainees to self-reflect on their performance. Archer's model for effective feedback includes: *self-monitoring* (reflection on action) supported by *external feedback* and linkage *with personal goals* (action plan) in a coherent process rather than a series of unrelated events.

Since formative assessment of clinical performance often includes feedback provided by an expert trainer after direct observation of a trainee at work[13], several assessment instruments have been proposed to enhance the effectiveness of this type of feedback, mostly based on the mini-CEX[14,15]. Currently, a number of comparable instruments are widely used in workplace-based assessment. In order to determine the occurrence of self-assessment, recommendations by the trainer and explicit formulation of an action plan - elements resonating with Archer's principles of *reflection, feedback* and *linking with personal goals*[12] -, Holmboe et al.[16] studied videotaped sessions in which supervisors provided oral feedback as part of a mini-CEX. Self-assessment *(reflection)* was found to be less frequent than recommendations made by supervisors *(feedback)*, while action plans *(linking with personal goals)* were rarely formulated. Based on these findings, we studied the effects on feedback of a modified mini-CEX. Like the original mini-CEX and similar assessment instruments[14], the instrument we studied is designed to generate feedback on observed performance during a clinical encounter. The instrument is tailored to practice settings in GP training in the Netherlands and the assessment form is designed to stimulate trainers and trainees to provide written narrative comments on trainee performance. We investigated the frequency of different types of comments invited in the form: self-reflection by the trainee, feedback from the trainer and an

Quality of written narrative feedback and reflection in a modified mini-clinical evaluation exercise: an observational study

Chapter 4

action plan proposed by both trainer and trainee. In view of the importance of the specificity of feedback[1,7,8,11,12], we also examined the specificity of the comments. We will use the word 'feedback' for written observations entered on the form by the trainer, 'reflection' for trainees' written self-assessments and 'action plan' for written descriptions of learning goals, plans to achieve them and methods to evaluate the outcome. We use 'comments' with reference to all kinds of text entered by trainees and trainers on the form, including 'feedback', 'reflections' and 'action plan'.

Method

Instrument

A modified mini-CEX was designed, including a form to evaluate trainees' competence during an observed clinical encounter in general practice with additional space provided for answers to questions inviting trainers to provide narrative feedback and trainees to provide narrative reflections on 'what went well' and 'what could have been done better', and for an action plan drawn up by trainer and trainee, comprising learning goals, steps for improvement and ways of evaluating these. The appendix presents an English translation of the Dutch form. As our study focused on the written narrative comments, we did not analyse the quantitative components of the assessment form (Appendix 1), comprising different aspects of three competencies (medical expert, communicator and professional) and an overall judgement of competence on a 10-point scale, as is customary in Dutch education[17].

Procedure and context

Within the postgraduate training programme in general practice in Nijmegen, the Netherlands, the above-described assessment form was introduced in March 2008 to stimulate GP trainers to give structured and systematic feedback on observed patient consultations conducted by GP trainees in the trainers' practices. Every three months, at least three such assessments must be conducted. With regard to the written comments, the instructions for using the form state that, after an observed consultation the GP trainee first should give a short reflection on his/ her performance, followed by feedback from the trainer, after which trainer and trainee use the reflection and the feedback to draw up an action plan to address weaknesses.

All trainers attended a training programme of half a day each month and two days annually, dealing with *all aspects* of the work of a GP trainer, including assessment, of which the modified mini-CEX is a part. All trainers received a basic introduction about the modified mini-CEX assessment form. They watched a video of a patient consultation, assessed it using the form and discussed this with one another. During the other parts of the programme, trainers had opportunities to discuss and

ask questions about observation and the use of the assessment form. *Trainees* were instructed about the overall assessment plan and the use of several assessment instruments (including the modified mini-CEX) at the beginning of their training. *Trainers and trainees* had permanent access to online manual providing information about the relevance of observation and written narrative feedback for educational purposes and about the procedures.

Participants and procedure

The above-described assessment form is in use during the postgraduate programme in general practice in Nijmegen, the Netherlands. During the first and last year of the three-year Dutch general practice programme, trainees work in a general practice, while training in the second year takes place in hospitals and other health care institutions. Since the assessment form is only used during the years in general practice, we studied the effects among first and third year trainees. Between March 2009 and September 2009 we asked GP trainees in Nijmegen who had started the first or third year of training in March of that year (N=69) to hand in their assessment forms. Since each trainee is supervised by one GP trainer for a whole year and each trainee is supervised by a different trainer, trainers and trainees were included in the study in pairs.

Participation was voluntarily. Trainees were informed of the purpose of the study and they could voluntarily hand in their assessment forms at the institution in Nijmegen. They could make their forms anonymous by using a unique number to code them. At the time of the data collection, no ethical review board for medical educational research existed in the Netherlands. We fully complied with ethical rules in terms of voluntariness and anonymity. The researchers had no hierarchical relationship with either the trainees or the trainers.

Data analysis

We first calculated the percentage of forms with written comments in response to the seven requests for comments in the form (1: reflection, what went well, 2: reflection, what could have been done better, 3: feedback, what went well, 4: feedback, what could have been done better, 5: action plan, learning goals, 6: action plan, plan, 7: action plan, method of evaluation) (Appendix 1). Next, we rated the specificity of the comments on a three-point scale (specific, moderately specific, not specific). Feedback and reflection were rated as specific when it was clear to *which part* of the consultation they related, *what* did and did not go well and/or *why* it did or did not go well. An example of a specific comment relating to 'what could have been done better' is: 'the consultation could have been finished more quickly'. A comment was rated as 'moderately specific' when it *only* indicated *which part* of the consultation did or did not go well or *what* did or did not go well or *why* a comment

was made. An example of a moderately specific comment on 'what went well' or 'what could have been done better' is: 'physical examination'. A comment was rated as 'not specific' when it was too general, relating to the consultation as a whole without specifying *which part* of the consultation was involved, *what* the comment referred to or *why* it was made. This type of a very general unspecified comment – such as 'pleasant contact' - does not seem very useful, especially when it is read after a period of time has elapsed, because by then it will be difficult to recall which aspects of the trainee's performance prompted the comment.

Comments on learning points were rated as specific if they explicitly stated *what* aspects needed more work. For example: 'exploration of the differential diagnosis'. A moderately specific comment is: 'continue to think critically and logically'. And a 'non-specific' comment: 'do more'. Comments about the planning were rated as specific if it was stated *how* the trainee could address a learning point. For example: 'reminder on my desk'. A moderately specific comment is: 'do not try to implement all learning points at once' and a non-specific comment: 'practise'. Comments about the evaluation were rated as specific if it was stated *how progress would be monitored,* for example: 'video recording'. A moderately specific comment on evaluation is: 'mutual assessment'.

The criteria for specific, moderately specific and non-specific comments were developed in a four step procedure. EP first read all the forms to gain an impression of how trainees had used the assessment form. Next, HM, AK and EP examined two forms (one with detailed and one with broad comments) and determined criteria for 'specificity', testing these criteria by independently rating five forms. After some small adjustments were made, the second version of the criteria was tested on 20 forms through independent rating by EP and AK (10 forms) and EP and HM (10 forms). The kappa coefficients for inter-coder agreement were .67 (EP/AK) and .77 (EP/HM). When agreement on coding was considered satisfactory, EP coded all the remaining forms. Discussion between EP and AK resolved uncertainty in regard of the rating for 30 of 485 forms due to poor legibility (14 forms) and doubts about categorisation (16 forms).

We used the data from trainees who handed in three or more forms to examine possible differences in specificity of comments of different trainee-trainer pairs. We calculated for each pair the percentage of specific comments, and we analysed differences (standard deviation) between pairs in specificity of comments for each of the seven questions in the form. For this calculation we dichotomised the results in 'specific comments' and a second category containing all other comments and blank forms.

Results

Of 485 forms returned by trainees, nine could not be related to an individual trainee, and the remaining 476 were from 54 different trainees, who had completed a mean number of 8.8 forms (SD 5.6; range 1-23). These trainees represented 78% of all the trainees invited to hand in their forms. Of the participating trainees, 68% were female and of the trainers 65% were male. These percentages are representative of the overall population of GP trainees and GP trainers in the Netherlands. The number of first year trainees exceeded that of third year trainees at the time of the study (57% were first year trainees). Also the first-year trainees returned more forms (78% first year trainees). Because of the between-group differences in return rate we examined whether there were quantitative differences between forms from first and third year trainees, but a chi-square analysis showed no significant differences (P > .05).

Table 1 shows the percentages of comments in response to the seven questions on the assessment form (Appendix 1), showing that reflection occurred less often than feedback and explicit formulation of an action plan was rare. Table 1 also shows the specificity of the comments for each of the seven questions. If comments were written down, the majority of the comments were specific (≥57%); and less than 10% was not specific.

Because of the differences between trainee-trainer pairs in the number of completed forms (range 1-23), we wanted to explore possible differences in specificity between comments of different trainee-trainer pairs. To examine this, we used the forms of trainee-trainer for which we had received at least three assessment forms. We calculated the mean percentage of specific comments per question per pair. Next we calculated the standard deviation (SD). Table 2 shows that SD's are high, which

	Forms with comments N (%)	Not specific %	Moderately specific %	Specific %
Reflection 'what went well'	259 (53.4)	9.3	33.6	57.1
Reflection 'what could be done better'	259 (53.4)	5.8	22.4	71.1
Feedback 'what went well'	433 (91.3)	5.9	26.9	67.3
Feedback 'what could be done better'	423 (87.2)	9.2	16.5	74.2
Action plan: 'Learning goal'	166 (34.2)	5.5	20.8	74.0
Action plan: 'Plan'	57 (11.8)	6.0	16.4	77.6
Action plan: 'Method of evaluation'	12 (2,5)	-	33.3	66.7

Table 1 Numbers and percentages of forms with specific types of comments and the specificity of the comments

	Mean percentage specific comments (SD)
Reflection 'what went well'	34,2 (29,7)
Reflection 'what could be done better'	42,7 (28,5)
Feedback 'what went well'	62,5 (22,7)
Feedback 'what could be done better'	67,5 (21,8)
Action plan: 'Learning goals'	27,0 (27,8)
Action plan: 'Plan'	11,2 (17,0)
Action plan: 'Method of evaluation'	2,3 (8,1)

Table 2 Percentage specific comments per pair (SD) N=50

means that there were large differences between pairs in the extent to which they formulated specific comments. Some pairs consistently provided specific comments on a certain question on all their assessment forms, while other pairs provided no specific comments relating to that question. This applies to all questions on the form, except for 'evaluation', which generally went unanswered.

Discussion

The results of this study show that the modified assessment form and procedure resulted in frequent reporting of feedback, less frequent reporting of reflection and only rare reporting of action plans. The results also show, however, that, generally, the reflections, feedback and action plans that were provided were specific. Based on the importance of specific comments as indicator of quality[1,7,8,11,12], we can conclude that, *if* comments were made, the modified assessment form elicited useful qualitative comments. It would be interesting, however, to further investigate the different frequencies of the different types of comments. Perhaps the modification, consisting only of encouragement and facilitation of written narrative reflections, feedback and an action plan, was not sufficiently powerful to induce trainers and trainees to make full use of the form. The way the assessment form was introduced and the availability of the online manual appear to have been inadequate to achieve reflective behaviour for all trainees and formulation of an action plan on a significant proportion of forms. Although the layout of the form directs which type of comments should be provided in which space, users retain the possibility to use it otherwise. It is also possible that feedback from the trainer and – to a lesser degree – reflection by the trainee are more firmly embedded in the assessment routine than linking these comments to the broader learning context by formulating an action plan. Perhaps, among trainers and trainees there already was a culture of giving feedback and, to a lesser extent of

reflecting on performance, but not (yet) of making plans for action to follow-up on feedback. This conclusion appears to be supported also by the large differences we found between trainee-trainer pairs. It seems that some pairs *do have* a culture of feedback, reflection and action plans, while for others such a culture remains to be developed. Apparently, some trainers and trainees do apply the information from assessment training and the online manual. These findings suggest that in order to enhance the effectiveness of assessment training, there should be a specific focus on reflection and action plans. Additionally, trainers and trainees who use all the feedback modalities might be asked to share their experiences.

An important quality of our study is the response rate. A large number of assessment forms was analysed and almost 80% of GP trainees in the sample handed in their forms. However, the number of completed forms per trainee differed widely, with some trainees handing in only one form, even though the minimum required for the study period was six. The overrepresentation of forms from first year trainees may be attributable to the introduction of the new assessment form. Since the version of the form that was the subject of this study was first introduced in March 2008, first year trainees had used the form from their first day of training, while for the third year trainees (who had started their training before 2008) it meant a change. However, since the percentages of written comments in response to the seven questions did not differ between these groups, the overrepresentation of first year residents apparently did not impact on the results.

It should be noted, that although the reported action plans were specific, this finding is based on limited data, because the majority of forms did not contain an action plan. Only a few trainee-trainer pairs provided comments relating to an action plan. Next, we only studied written narrative comments entered in the assessment form. This is a limitation because we do not know what actually happened during the discussion between trainee and trainer when the text was formulated, and therefore a comparison with the results of Holmboe et al. cannot be made[16]. We chose our method because written narrative texts are one of the positive qualities of these formative assessment forms. Forms can be stored by trainees in their portfolio to help them reflect on a series of mini-CEX results in order to formulate learning goals, and they can help trainees and trainers to gain an overall impression of development of performance. Another limitation is the focus on the qualitative part of the assessment form. Further research should examine relationships between narrative feedback and the quantitative part of the assessment form.

In this study we looked at the written results of workplace-based observation and feedback. In a qualitative study[18] we examined ideas, barriers and motives experienced by trainers and trainees in relation to observation, reflection and feedback. Based on the results of the present study we would recommend a different approach to training to stimulate reflection in trainees and more attention to the formulation of an action

plan, elements that are important for the effectiveness of feedback[12,19]. Further research is needed to explore how feedback and reflections that are specific as well as goal-oriented, as evidence by the formulation of an action plan, impact on performance improvement, the ultimate aim of assessment of observed performance. The implications of the substantial differences between trainee-trainer pairs in relation to the percentages of specific comments require further investigation as well.

Conclusion

The main findings of this study are that self-reflection by the trainee and formulation of an action plan were not uniformly reported on the assessment forms, the comments on the forms were generally specific and there were substantial, consistent differences between trainee-trainer pairs in the provision of specific comments. Based on these findings, we conclude that it is not so much the instrument (form and instructions) but rather the users that determine how the modified mini-CEX form is used. This suggests that intervention to improve the educational effectiveness of the feedback procedure should focus more on the users than on the instruments.

References

1. Black, P. & William, D. (1998). Assessment and classroom learning. *Assessment in Education, 5,* 7 – 74.
2. Black, P. & William, D. (1998). Inside the black box – raising standards through classroom assessment. *Phi Delta Kappa, 80,* 139 – 149.
3. Norcini, J.B. & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher, 29,* 855 – 871.
4. Yorke, M. (2003). Formative assessment in higher education: moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45,* 477 – 501.
5. Krasne, S.W., Wimmers, P.F., Relan, A. & Drake, T.A. (2006). Differential effects of two types of formative assessment in predicting performance of first-year medical students. *Advances in Health Sciences Education, 11,* 155 – 171.
6. Rudolph, J.W., Simon, R., Reamer, D.B. & Eppich W.J. (2008). Debriefing as formative assessment: closing performance gaps in medical education. *Academic Emergency Medicine, 15,* 1010 – 1016.
7. Rolfe, I. & McPherson, J. (1995). Formative assessment – how am I doing? *Lancet, 345,* 837 – 839.
8. Shute, V.J. (2008). Focus on formative feedback. *Review of Educational Research, 78,* 153 – 189.
9. Overeem, K., Lombarts, M.J., Arah, O.A., Klazinga, N.S., Grol, R.P. & Wollersheim, H.C. (2010). Three methods of multi-source feedback compared: a plea for narrative comments on coworkers' perspectives. *Medical Teacher, 32,* 141 – 147.
10. Govaerts, M.J., Vleuten, van der, C.P., Schuwirth, L.W. & Muijtjens, A.M. (2005). The use of observational diaries in in-training evaluation: student perceptions. *Advances in Health Sciences Education, 10,* 171 – 188.
11. Sargeant, J., Mann, K., Vleuten, van der, C. & Metsemakers, J. (2008). 'Directed' self-assessment: practice and feedback within a social context. *Journal of Continuing Education in the Health Professions, 28,* 47 – 54.
12. Archer, J.C. (2010). State of the science in health professional education: effective feedback. *Medical Education, 44,* 101 – 108.
13. Watling, C.J. & Lingard, L. (2012). Toward meaningful evaluation of medical trainees: the influence of participants' perceptions of the process. *Advances in Health Sciences Education, 17,* 183 – 194.
14. Pelgrim, E.A., Kramer, A.W., Mokkink, H.G., Elsen, van den, L., Grol, R.P. & Vleuten, van der, C.P. (2011). In-training assessment using direct observation of single-patient encounters: a literature review. *Advance in Health Sciences Education, 16,* 131 – 142.
15. Kogan, J.R., Holmboe, E.S. & Hauer, K.E. (2009). Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA, 302,* 1316 – 1326.
16. Holmboe, E.S., Yepes, M., Williams, F. & Huot, S.J. (2004). Feedback and the mini clinical evaluation exercise. *Journal of General Internal Medicine, 19,* 558 – 561.
17. Cate, ten, J., Braak, E., Fenkel, J. & Pol, van der, A. (2006). De 4- tot 10 verwacht niveau-schaal (410VN-schaal) bij persoonlijke beoordelingen. *Tijdschrift voor Medisch Onderwijs, 25,* 31.
18. Pelgrim, E.A.M., Kramer, A.W.M., Mokkink, H.G.A. & Vleuten, van der, C.P.M. (2012). The process of feedback in workplace-based assessment: organization, delivery, continuity. *Medical Education, 46,* 604 – 612.
19. Sargeant, J.M., Mann, K.V., Vleuten, van der, C.P. & Metsemakers, J.F. (2009). Reflection: a link between receiving and using assessment feedback. *Advances in Health Sciences Education, 14,* 399 – 410.

98

99

Quality of written narrative feedback and reflection in a modified mini-clinical evaluation exercise: an observational study

Chapter 4

Short Clinical Assessment

Name of GP-resident:
    Name of GP-trainer:
    Date:

Setting: consultation / telephone / home-call / duty / other, namely...

Patient: new / follow-up   **complexity**: low / moderate / high

1. MEDICAL EXPERT	≤4 5 6 7 8 9 10 not observed

- Asks for clarification if necessary (medical history, psychosocial)
- Asks relevant questions
- Performs relevant physical examination in a correct manner
- Establishes correct (preliminary) diagnoses
- Offers adequate policy/ help
- Avoids redundant medical care
- Logical sequencing of contact

2. COMMUNICATOR	≤4 5 6 7 8 9 10 not observed

- Clarifies request for help
- Asks for cognitions, emotions and illness behaviour
- Links the analyses and the request for help
- Sets the management plan together with the patient
- Appropriately tunes language
- Explains understandable
- Empathizes
- Leads and is timely

3. PROFESSIONAL	≤4 5 6 7 8 9 10 not observed

- Treats respectfully and is sensitive
- Takes responsibility
- Balances involvement and distance

4. OVERALL JUDGEMENT OF COMPETENCE	≤4 5 6 7 8 9 10

REFLECTION OF GP-RESIDENT	**TRAINERS' FEEDBACK**
What went well?	What went well?
What could be done better?	What could be done better?

ACTION PLAN *(specify: learning goals, plan, and method of evaluation)*

Appendix 1 The adjusted mini-CEX tool

Chapter 5

Reflection as a component of formative assessment appears to be instrumental in promoting the use of feedback; an observational study

E.A.M. Pelgrim, A.W.M. Kramer, H.G.A. Mokkink, C.P.M. van der Vleuten

Abstract

Background: Although the literature suggests that reflection has a positive impact on learning, there is a paucity of evidence to support this notion.

Aim: We investigated feedback and reflection in relation to the likelihood that feedback will be used to inform action plans. We hypothesised that feedback and reflection present a cumulative sequence (i.e. trainers only pay attention to trainees' reflections when they provided specific feedback) and we hypothesised a supplementary effect of reflection.

Method: We analysed copies of assessment forms containing trainees' reflections and trainers' feedback on observed clinical performance. We determined whether the response patterns revealed cumulative sequences in line with the Guttman scale. We further examined the relationship between reflection, feedback and the mean number of specific comments related to an action plan (ANOVA) and we calculated two effect sizes.

Results: Both hypotheses were confirmed by the results. The response pattern found showed an almost perfect fit with the Guttman scale (0.99) and reflection seems to have supplementary effect on the variable action plan.

Conclusions: Reflection only occurs when a trainer has provided specific feedback; trainees who reflect on their performance are more likely to make use of feedback. These results confirm findings and suggestions reported in the literature.

Introduction

The aim of this study was to clarify the relationship between feedback, reflection and the use of feedback. Research on formative assessment and feedback suggests that these are powerful tools to change trainees' behaviour[1-4]. In this paper we use the following definition of formative assessment: "an instructional intervention aimed at evaluating performance and providing feedback on trainees' strengths and weaknesses"[1,5]. Several studies have shown limited benefit from feedback in the form of numerical marks[1, 6-8] and it is suggested that quantitative feedback should be more detailed and feedback preferably narrative, explicating in which areas more work is needed. Positive effects of narrative feedback have been reported by Overeem et al.[9], who found higher satisfaction with this type of feedback, and by Govaerts et al.[10], who suggested that it can improve in-training evaluation. Both Archer[11] and Sargeant[12] claim that specific feedback is more readily assimilated, while Shute[8] found a significant positive influence of specific feedback on performance.

Reflection is an evaluative activity which, similar to feedback, has a potentially powerful learning effect, both in educational settings and in today's complex and changing health care systems, by giving meaning to complex situations and enabling deeper learning from experiences[13]. In a similar vein, Sargeant et al.[14] have described reflection and reflective practice as essential attributes of a competent health care professional[1].

Two 'objects' of reflection in formative assessment can be distinguished when analysing literature: reflection on performance and reflection on feedback. Feedback recipients engage in reflection *on performance* to self-monitor their performance prior to receiving external feedback[11,13], while after receiving feedback they engage in reflection *on feedback received* to form an opinion of that feedback. Both objects of reflection are part of a complex situation, wherein reflective practice is a way to learn and give meaning. Sargeant et al.[14] concluded that reflection on feedback supported its acceptance and use for improvement. This conclusion resulted from an interview study among general practitioners participating in a multi-source feedback programme, where reflection-on-feedback appeared to be instrumental in decisions to accept and use feedback. Although there is evidence supporting the effects of reflection-on-feedback, there is no empirical evidence for the impact of reflection-on-performance on learning and feedback use. In this study we focused on reflection-on-performance by exploring how it relates to feedback and feedback use.

Despite the absence of empirical evidence, the literature suggests that reflection-on-performance may have a positive effect on the use of feedback. Archer highlights that feedback can be truly effective only if it is built on self-monitoring informed by external feedback, thereby suggesting that reflection (on-performance) has a positive effect on learning[11]. This conclusion was based on a summary of key areas of the

feedback literature and a model based on this summary. Earlier, Pendleton developed the so-called Pendleton Rules[15] representing recommendations for a 'good way of giving feedback', including the recommendation that after trainee performance, first the trainee and after that the trainer should comment on what went well and why and what could have been done differently. Although Pendleton stresses the importance of reflection, the Rules were primarily developed to counteract the observed effect in medical education that feedback was traditionally negative, pointing out trainees' errors while failing to draw attention to their strengths[16]. Pendleton, however, did not actually examine the effects of reflection on feedback use. In an earlier empirical study on formative assessment and feedback using an adapted mini-CEX instrument with a form inviting trainers to provide narrative feedback and trainees to provide narrative reflections, we found that feedback occurred more frequently than reflections[17].

In the present study we examined the relationship between feedback, reflection and feedback use by analysing completed forms of the above-described instrument. We hypothesised that feedback and reflection present a cumulative sequence. In other words, whether trainers pay attention to trainees' reflections depends on whether they provide specific feedback. Here a distinction must be made between the sequence during the feedback process and the occurrence of reflection and feedback on the assessment forms. The sequence during the feedback process is line with the Pendleton rules (reflection – feedback). Our hypothesis is that on the assessment form specific reflection is only written down when specific feedback is given (cumulative sequence of feedback – reflection). Most trainers are quite used to giving specific feedback but less used to paying attention to trainees' reflections, which is a more recent development in medical education. Based on the literature[11, 14-16] we also hypothesised that reflection has a supplementary effect on the use of feedback, with trainees who reflect on performance being more likely to make use of the feedback they receive.

Methods

Design
In a cross-sectional study, we examined formative assessment forms (Appendix 1) completed by general practice (GP) trainees and their trainers after the trainer had observed the trainee conducting a consultation with a real patient. The form requests trainees to enter reflections on their performance, trainers to enter narrative feedback and trainer and trainee to formulate a joint action plan to achieve improvement guided by the feedback and the reflections. We first examined if specific feedback and specific reflections showed the expected cumulative pattern of

reflection depending on the presence of specific feedback. Secondly, we tested the hypothesis that better, i.e. more specific, feedback and reflection stimulate the use of feedback operationalised as the formulation of an action plan.

Participants
The study was conducted among GP trainees and their trainers in the three-year postgraduate GP training programme in Nijmegen, the Netherlands. Trainees work in a general practice supervised by a GP trainer during the first and third years of training and in the second year they work in hospitals and other health care institutions. The data for the present study were collected from forms used by trainees during the first and third years of training. Between March 2009 and September 2009 trainees who had started their first or third year of training in March 2009 (N=69) were invited to hand in copies of completed assessment forms. Since trainees are supervised by one GP trainer for the duration of a whole year, the trainer-trainee pairs completing the forms did not change in the course of the study, and study participants were therefore defined as trainer-trainee pairs.

Trainees were informed of the purpose of the study and invited to hand in their assessment forms at the institute of family medicine in Nijmegen. Participation was voluntary and anonymous, and a unique code number could be used to make forms anonymous. At the time of the data collection, there was no ethical review board for medical educational research in the Netherlands, while hospital review boards did not consider it to be their task to evaluate this type of research. However, we fully complied with ethical rules in terms of voluntariness and anonymity.

Instrument and setting
The assessment instrument used in this study was adapted from the traditional mini-CEX form by providing separate spaces for trainers to write narrative feedback, for trainees to enter narrative reflections and for an action plan, developed jointly by trainer and trainee, describing how they planned to use the feedback for learning and performance improvement (Appendix 1). As recommended in the Pendleton rules[15], the instructions on how to use the form requested trainees to describe their reflections and trainers to give feedback by answering the questions 'what went well' and 'what could have been done better'. Trainers and trainees were additionally asked to describe learning goals and a plan of action and how achievement of these would be evaluated. The form also comprised other feedback, but in this study we focused exclusively on the narrative comments entered in the forms (Appendix 1).

In the Nijmegen postgraduate training programme the assessment form is routinely used to elicit structured and systematic formative feedback on consultations with real patients conducted by a trainee and directly observed by the trainer. The training institute recommends that this type of assessment should take place at least

monthly. At the beginning of the year (March 2009), both trainers and trainees were instructed about the form and the feedback procedure, and after that they had continuous access to an online instruction manual. Trainers and trainees were instructed that, after an observed consultation, the trainee should first give a short reflection on his/her performance, followed by feedback from the trainer after which an action plan should be developed by trainer and trainee to address specific issues raised during the procedure. The assessment form could be used also for assessment of trainers by trainees and assessment of trainees by another observer, another trainee for instance. Trainees kept the completed forms in their portfolio and handed in a (anonymous) copy for the study.

Data analysis
- *The assessment of forms*

In an earlier study[17], we analysed completed forms to determine whether feedback, reflections and action plan were specific. Three researchers (EP, AK and HM) performed this analysis. The kappa coefficients for inter-coder agreement were .67 (EP/AK) and .77 (EP/HM). Reflections and feedback were rated as specific when it was clearly stated to w*hich part* of the consultation they related, what did and did not go well and/or *why*. An example of a specific comment is: 'the consultation could have been finished more quickly'. Action plans were rated as specific if one or more learning objectives, plans or evaluation methods were described. An example of a specific action plan is: 'More practice in inserting a urine catheter'.

- *Aggregation and calculation of variables*

Since the participants in the study were trainer-trainee pairs, the data were aggregated for these pairs. We analysed only forms completed by trainer-trainee pairs for which we had received more than three forms. For each pair, we calculated the percentage of forms containing specific feedback and specific reflections, respectively. The variables feedback and reflections were rated separately, and characterised as specific if at least 50% of the forms of a trainer-trainee pair contained specific feedback and specific reflections, respectively. This resulted in four types of trainer-trainee pairs: 'no specific feedback and no specific reflections', 'specific feedback and no specific reflections', 'no specific feedback and specific reflections' and 'specific feedback and specific reflections'. In the analysis of the variable 'action plan' we calculated for each pair the mean number of specific comments relating to an action plan.

- *Guttman scale*

To test our first hypothesis (cumulative sequence of specific feedback and specific reflections) we examined whether the data revealed response patterns that were consistent with the three patterns postulated by the Guttman scale[18]: 'no specific feedback and no specific reflections' (the poorest pattern), 'specific feedback and no specific reflections' and 'specific feedback and specific reflections' (the best pattern). Note that the scalogram model postulates that the pattern 'specific reflections and no specific feedback' does not occur. Coefficients for the Guttman scale (percentage of scores that fit with the model) were calculated. These coefficients are considered acceptable at a value of .90 or higher[18].

- *Difference of means and effect size*

To examine the hypothesised supplementary effect of reflection we calculated the mean number of specific comments related to an action plan and analysed the relationship between feedback and reflection patterns (ANOVA) and calculated two effect sizes. We compared pairs with 'specific feedback and no specific reflections' and pairs with 'specific feedback and specific reflections', using as reference group pairs with 'no specific feedback and no specific reflections'. The effect sizes compared the means of the two independent groups[19]. Effect sizes of .20, .50 and .80 or higher were considered to be small, moderate and large, respectively[19].

Results

Of 485 forms returned by trainees, nine forms could not be traced to an individual trainee; the remaining 476 forms were from 54 different trainees, implying that 78% of trainees returned one or more forms. Of these trainees, 68% were female and of the trainers 65% were male. These percentages are representative of the overall population of GP trainees and GP trainers in the Netherlands. Trainees handed in a mean number of completed forms of 8.8 (SD 5.6; range 1-23). Of the 476 assessment forms, 69 related to a consultation observed by another trainee or to evaluation of the trainer. Elimination of these forms left a total of 407 forms, from which we selected forms of trainer-trainee pairs for which at least three completed forms were available. This resulted in a total of 400 forms from fifty trainer-trainee pairs (Table 1).

The analysis showed that 66% and 34% of the forms contained specific feedback and specific reflections, respectively. The forms contained an average of 0.53 specific comments related to an action plan, which is slightly more than one action plan per two forms.

Number of forms	Number of pairs
1	1
2	3
3	9
4	5
5	1
6	7
7	3
8	5
9	2
10	6
11	4
12	3
14	1
16	1
17	2
23	1
Total 407	Total 54

Table 1 Number of forms per trainer-trainee pair

The process of feedback and reflection (Guttman scale)
Only one pair showed an inconsistent pattern (specific reflections and no specific feedback) (Table 2). The other pairs were evenly distributed over the consistent patterns, resulting in an almost perfect Guttman scale with a high coefficient (0.99).

	Frequency
No specific feedback and no specific reflection	16 (32%)
Specific feedback and no specific reflection	17 (34%)
No specific feedback and specific reflection	1 (2%)
Specific feedback and specific reflection	16 (32%)

Table 2 Number of pairs by comments on feedback and reflection (N=50)

The effect of feedback and reflection on action plan (effect size)
Pairs with 'no specific feedback and no specific reflection' (reference) had 0.25 comments per form (or one comment per four forms). Pairs with 'specific feedback and no specific reflections' had 0.33 comments per form (or one comment per three forms). This suggests a moderate effect of feedback on action plan. Pairs with the best pattern, i.e. 'specific feedback and specific reflections', had 1.02 comments per form, which is considered to be a large effect (Table 3).

	Frequency	Specific comments on action plan	Effect size
No specific feedback and no specific reflection	16	0.25	Reference
Specific feedback and no specific reflection	17	0.33	0.21
Specific feedback and specific reflection	16	1.02	1.71

Table 3 Mean number of specific comments on action plan, and effect size

Discussion

The results support both hypotheses. We found a cumulative sequence of feedback and reflections as well as a supplementary effect of reflections on the use of feedback. Specific reflections were found only in combination with specific feedback and they occurred more often in combination with specific action plans. Our results are not only in line with suggestions in the literature that reflection-on-performance has a positive effect on feedback use[11] but they additionally suggest that the effect we found for reflection-on-performance is comparable to the effect of reflection-on-feedback reported by Sargeant et al.[14].

The results provide support for the existence of a hierarchy in the quality of feedback. Some trainer-trainee pairs showed neither specific feedback nor specific reflections, some pairs showed only specific feedback and some pairs showed both specific feedback and specific reflections. Pairs with specific feedback and specific reflections appeared to be more inclined to also formulate a specific action plan. These results offer additional support for the three-step model we developed in an earlier study[20] and reflecting the findings that feedback in workplace-based assessment appears to be of greater benefit to trainees if: (1) observation and feedback are planned by trainee and trainer; (2) the content and delivery of the feedback are adequate; (3) trainees use feedback to guide their learning by linking it to learning goals. These steps are considered to be sequential, with steps 2 and 3

relating to the delivery of feedback and its use for further learning, respectively. The results of the present study add to this model by showing a relationship between better delivery of feedback (i.e. in accordance with descriptions in the literature[11,15] and incorporating reflection-on-performance) and more frequent occurrence of step 3 of the model[20].

The sequential process of feedback–reflection–action plan found in this study might be used in faculty development in the area of feedback. Feedback has received considerable attention in (medical) education for quite some time, and trainers have learned a great deal about the importance of feedback and how it is best provided. The importance of reflection and its connection with the use of feedback in setting learning goals is a later development, however, and these notions have not yet found their way into the standard feedback procedure of many trainers. It seems therefore advisable that the role of reflection in the feedback process and its connection with learning goals should be included in faculty development activities aimed at improving trainers' feedback skills.

A possible explanation for the influence of reflection on feedback use is that the independent variable 'formulation of an action plan' seems to bear a close resemblance to reflection. Similar to reflection-on-performance, formulating an action plan could be interpreted as a way to give meaning to the situation where performance takes place. For trainees who habitually reflect on their performance, it is probably only a small step to elaborate on and use insights obtained from reflection as starting points for further learning. The notion that reflective practice can directly promote the formulation of action plans underscores our recommendation that faculty development activities should pay attention to the deployment of reflective practice in the feedback process. It could also be the case that there are trainer-trainee pairs who are more willing to follow the instructions of the GP-training institute and therefore use the whole assessment form. Not because they are more willing to reflect, but just to comply with the instructions. In that case, the question is whether these trainees learn better, or learn more than those who do not comply with instructions. However, there remains a need for further research into feedback use. In the present study we operationalised feedback use as 'formulation of an action plan', but this is only a proxy for feedback use. Further research should also investigate adherence to action plans. Our data did not allow investigation of questions such as: do trainees comply with their plans and do trainees who formulate an action plan learn more compared to trainees who do not formulate an action plan.

We included only a limited sample of fifty trainer-trainee pairs from one training institute who participated voluntarily. Nevertheless, the response rate was acceptable and the participating pairs provided 400 assessment forms for analysis, which seems to offer a sufficiently solid basis for the results. Another limitation is that we studied only the written comments entered in the forms. This means that we have no

information about what actually occurred when trainees and trainers discussed trainee performance. We used this method of analysis because the option to enter written comments was a special feature added to the form to enhance the quality of the feedback. The forms are kept by trainees in their portfolios and can help them to develop relevant learning goals based on a series of assessments while at the same time giving trainees and trainers an overall impression of trainee progress.

In conclusion, we think that our study adds to the existing insights related to reflection and feedback in formative assessment. Our results are in line with evidence from the literature that feedback can offer new information and that reflection-on-feedback facilitates the acceptance of feedback[14] and they also support theory on the usefulness of reflection-on-performance[11] by offering new evidence that reflection-on-performance can promote feedback use. Apparently, a reflective attitude towards one's own performance plays an important role in obtaining and reflecting on external feedback and in applying it to enhance learning and improve performance. We think that our results warrant the recommendation that reflection should be included in faculty development activities to help trainers optimise the feedback process. Based on our results, we expect that this can stimulate the use of feedback. Further research should focus on the actual use of feedback and how it affects doctors' performance in clinical practice.

References

1. Black, P. & William, D. (1998) Inside the black box – raising standards through classroom assessment. *Phi Delta Kappa, 80,* 139 – 149.
2. Black, P. & William, D. (1998). Assessment and classroom learning. *Assessment in Education, 5,* 7 – 74.
3. Norcini, J.B. & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher, 29,* 855 – 871.
4. Yorke, M. (2003). Formative assessment in higher education: moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45,* 477 – 501.
5. Krasne, S.W., Wimmers, P.F., Relan, A. & Drake, T.A. (2006). Differential effects of two types of formative assessment in predicting performance of first-year medical students. *Advances in Health Sciences Education, 11,* 155 – 171.
6. Rolfe, I. & McPherson, J. (1995). Formative assessment – how am I doing? *Lancet, 345,* 837 – 839.
7. Vleuten, van der, C.P., Schuwirth, L.W., Scheele, F., Driessen, E.W. & Hodges, B. (2010). The assessment of professional competence: building blocks for theory development. *Best Practice and Research Clinical Obstetrics and Gynaecology, 24,* 703 – 719.
8. Shute, V.J. (2008). Focus on formative feedback. *Review of Educational Research, 78,* 153 – 189.
9. Overeem, K., Lombarts, M.J., Arah, O.A., Klazinga, N.S., Grol, R.P. & Wollersheim, H.C. (2010). Three methods of multi-source feedback compared: a plea for narrative comments on coworkers' perspectives. *Medical Teacher, 32,* 141 – 147.
10. Govaerts, M.J., Vleuten, van der, C.P., Schuwirth, L.W. & Muijtjens, A.M. (2005). The use of observational diaries in in-training evaluation: student perceptions. *Advances in Health Sciences Education, 10,* 171 – 188.
11. Archer, J.C. (2010). State of the science in health professional education: effective feedback. *Medical Education, 44,* 101 – 108.
12. Sargeant, J., Mann, K., Vleuten, van der, C. & Metsemakers, J. (2008). 'Directed' self-assessment: practice and feedback within a social context. *Journal of Continuing Education in the Health Professions, 28,* 47 – 54.
13. Mann, K., Gordon, J. & MacLeod, A. (2009). Reflection and reflective practice in health professions education: a systematic review. *Advances in Health Sciences Education, 14,* 595 – 621.
14. Sargeant, J.M., Mann, K.V., Vleuten, van der, C.P. & Metsemakers, J.F. (2009). Reflection: a link between receiving and using assessment feedback. *Advances in Health Sciences Education, 14,* 399 – 410.
15. Pendelton, D.S., Schofield, T., Tate, P. & Havelock, P. (1984). *The consultation: an approach to learning and teaching.* Oxford, Oxford University Press.
16. Swanwick, T. (2010). *Understanding medical education: evidence, theory and practice.* Oxford, Wiley - Blackwell.
17. Pelgrim, E.A.M., Kramer, A.W.M., Mokkink, H.G.A. & Vleuten, van der, C.P.M. (2012). Quality of written narrative feedback and reflection in a modified mini-clinical evaluation exercise: an observational study. *BMC Medical Education, 12,* 97.
18. Hays, R.D. & Ellickson, P.L. (1990). Longitudinal scalogram analysis – a methodology and microcomputer program for Guttman scale analysis of longitudinal data. *Behavior Research Methods, 22,* 162 – 166.
19. Hojat, M. & Xu, G. (2004). A visitor's guide to effect sizes – Statistical significance versus practical (clinical) importance of research findings. *Advances in Health Sciences Education, 9,* 241 – 249.
20. Pelgrim, E.A.M., Kramer, A.W.M., Mokkink, H.G.A. & Vleuten, van der, C.P.M. (2012). The process of feedback in workplace-based assessment: organization, delivery, continuity. *Medical Education, 46,* 604 – 612.

Short Clinical Assessment

Name of GP-resident:
    Name of GP-trainer:
    Date:

Setting: consultation / telephone / home-call / duty / other, namely...

Patient: new / follow-up   **complexity**: low / moderate / high

1. MEDICAL EXPERT	≤ 4 5 6 7 8 9 10 not observed

- Asks for clarification if necessary (medical history, psychosocial)
- Asks relevant questions
- Performs relevant physical examination in a correct manner
- Establishes correct (preliminary) diagnoses
- Offers adequate policy/ help
- Avoids redundant medical care
- Logical sequencing of contact

2. COMMUNICATOR	≤ 4 5 6 7 8 9 10 not observed

- Clarifies request for help
- Asks for cognitions, emotions and illness behaviour
- Links the analyses and the request for help
- Sets the management plan together with the patient
- Appropriately tunes language
- Explains understandable
- Empathizes
- Leads and is timely

3. PROFESSIONAL	≤ 4 5 6 7 8 9 10 not observed

- Treats respectfully and is sensitive
- Takes responsibility
- Balances involvement and distance

4. OVERALL JUDGEMENT OF COMPETENCE	≤ 4 5 6 7 8 9 10

REFLECTION OF GP-RESIDENT	**TRAINERS' FEEDBACK**
What went well?	What went well?
What could be done better?	What could be done better?

ACTION PLAN *(specify: learning goals, plan, and method of evaluation)*

Appendix 1 The adjusted mini-CEX tool

Chapter 6

Factors influencing trainers' feedback-giving behaviour

Abstract

Background The literature provides some insight into the role of feedback givers, but little information about factors influencing 'feedback-giving behaviours'. We looked for relationships between characteristics of feedback givers (self-efficacy, task perception, neuroticism, extraversion, agreeableness and conscientiousness) and elements of observation and feedback (frequency, quality of content and consequential impact).

Method We developed and tested several hypotheses regarding the characteristics and elements in a cross-sectional digital survey among GP trainers and their trainees. We conducted bivariate analysis using Pearson correlations and performed multiple regression analysis.

Results Sixty-two trainer-trainee couples from three Dutch institutions for postgraduate GP training participated in the study. Trainers' task perception and neurotic personality correlated positively with frequency of feedback and quality of feedback content. Multiple regression analysis supported positive correlations between task perception and frequency of feedback and between neuroticism and quality of feedback content. No other correlations were found.

Discussion This study contributes to the literature on feedback giving by revealing factors that influence feedback-giving behaviour, namely neuroticism and task perception. Trainers whose task perception included facilitation of observation and feedback (task perception) and trainers who were concerned about the safety of their patients during consultations with trainees (neuroticism) engaged more frequently in observation and feedback and gave feedback of higher quality.

Background

GP trainees in the Netherlands spend much of their postgraduate training working in a general practice where they are supervised by the same GP trainer for a prolonged period of time (around one year). The trainer can observe the trainee during patient consultations and give feedback on performance. Research has shown that learning from feedback depends on three elements: the feedback giver, the feedback recipient and the feedback, i.e. the message about performance. Two reviews on the *message*[1,2], each covering over one hundred articles and book chapters, showed that feedback is best aimed at a problem or task, at the process or at self-regulation but never at a personal trait of the recipient. Feedback should also enable comparison with an established standard. Psychological studies and recent medical education research[3-11] have shown that the role of the *feedback recipient* is an active one with recipients seeking information explicitly, for example by asking a trainer: 'how am I doing', or implicitly by making use of feedback intended for others[5,7,8]. Feedback-seeking behaviour is affected by personal and contextual factors[3,4,6] and feedback recipients decide whether to accept and use feedback[9], based on factors like self-reflection on performance[12], reflection on feedback[13] and the perceived credibility of the feedback source[9]. A literature search on the feedback context, the *feedback giver* in particular, however, yielded far fewer results than searches on the *message* and the *recipient.*

In the present study we focus on the feedback giver. In the only study we found on feedback-giving behaviour, Adams reported that feedback-giving behaviour in the US Army helicopter training school was influenced by a positive affect of trainers, with trainees who were well liked by trainers receiving less positive and less specific feedback. The author concluded therefore that popular trainees should be especially vigilant in pursuing feedback necessary for their personal development[14]. Evidence from studies on feedback seeking showed that benefits of feedback as perceived by trainees depended on the trainer[8,10,11]. Trainers who combined a supportive (involved and accessible) and instrumental (focusing on rules and responsibilities) supervisory style were more successful in convincing residents of the value of directly asking for feedback[8,10,11]. Additional evidence from an earlier study[15] showed that trainers who were active feedback givers were able to overcome trainees' feedback averse behaviours. Despite quite a few studies on the role of the feedback giver, not much is known about factors influencing 'feedback-giving behaviour', although an interview study by Kogan et al.[16] showed greater perceived ease of giving feedback among trainers who were more self-confident about feedback giving.

To contribute to the literature on feedback giving, we investigated the impact of personal characteristics of feedback givers on feedback-giving behaviour, specifically on feedback after observation of single patient encounters in postgraduate GP

training. This feedback process encompasses three elements: *organization* (frequency), the *quality of feedback content* (does the feedback adhere to directives about the 'message') and *consequential impact* (does the trainee use the feedback to determine and pursue learning goals and link present and previous feedback)[15]. Based on the literature on trainer effects on these three elements[12,15,17] and on the researchers' experiences, we hypothesized six correlations between trainer characteristics and elements of the feedback process.

Hypothesis 1.

Based on a study by Kogan et al. into effects of trainers' self-confidence[16], we formulated a hypothesis on self-efficacy, i.e. the belief in one's ability to succeed in a specific situation[18], in this case the 'preconditions' of the feedback process, i.e. arrangements to facilitate observation and feedback and support trainees in using feedback. We hypothesized that trainers with strong self-efficacy are more inclined to give feedback *(frequency),* give better feedback *(quality of content)* and are better able to convince trainees to use feedback for improvement *(consequential impact).*

> *Hypothesis 1: A trainer's high sense of self-efficacy (concerning preconditions for the feedback process) is positively correlated with the three elements of the feedback process, frequency, quality of content and consequential impact.*

Hypothesis 2.

Previous research showed that giving feedback is a core characteristic of competent trainers[19,20]. It seems plausible that trainers who do not consider it to be their task to create favourable preconditions for feedback should have a negative effect on the three elements of the feedback process. In line with this reasoning we proposed the following hypothesis:

> *Hypothesis 2: A trainer who sees it as his or her professional task to create positive preconditions for the feedback process (positive task perception) shows better feedback-giving behaviour in terms of frequency, quality of content and consequential impact.*

Hypotheses 3, 4, 5 and 6.

The Big Five refers to a taxonomy of personality traits comprising five domains: neuroticism, extraversion, conscientiousness, agreeableness and openness to experience[21,22]. Based on a study by Krasman of the impact of these traits on *feedback-seeking behaviour*[6], we formulated four hypothetical relationships between the Big Five and elements of *feedback-giving behaviour.*

Neuroticism refers to a person's emotional stability. Krasman demonstrated that neurotic people tend to seek more feedback, probably to alleviate their strong sense of insecurity[6]. A similar effect might be seen in trainers who give more feedback to compensate for their feelings of insecurity about entrusting the care of their patients to a trainee. Using a similar line of reasoning, we hypothesized that more neurotic trainers should be more inclined to make sure their trainees do use feedback (consequential impact). Neuroticism could have a negative effect on the content of feedback because neurotic people are easily frustrated, irritable and prone to react violently[22]. These considerations resulted in the following hypothesis:

> *Hypothesis 3: Compared to trainers with a fairly stable emotional make-up, trainers with a neurotic personality score higher on frequency and consequential impact of feedback and lower on the quality of the content of the feedback process.*

Extraverted people are very sociable, enthusiastic and action-oriented; they like to talk and be the centre of attention in groups. Although introverted people too may be very active and energetic, they are less sociable. Krasman demonstrated that extraverted people are inclined to seek more feedback[6]. Because feedback giving is an interpersonal activity, extraverted trainers may engage more intensely in feedback giving with a positive effect on frequency and consequential impact. We saw no reason to expect extraverted trainers to give better feedback (quality).

> *Hypothesis 4: The personal characteristic extraversion correlates positively with the frequency and consequential impact of the feedback process.*

Although people who rank high on *agreeableness* are interested in other people's concerns, Krasman found no correlation between agreeableness and feedback-seeking behaviour. Since feedback-giving behaviour is related to an interest in other people's concerns, we considered it nevertheless plausible that agreeableness should correlate positively with frequency, consequential impact and quality of feedback.

> *Hypothesis 5: The personal characteristic agreeableness is positively correlated with the elements frequency, quality and consequential impact of the feedback process.*

People with a *conscientious* personality tend to prefer planned to spontaneous behaviour, have strong self-discipline and a strong sense of duty and aim for achievement against certain standards or outside expectations. Conscientious people are intent on performing their tasks properly and Krasman found that they sought more feedback[6]. It therefore seemed plausible that conscientious trainers should score high on the three elements of feedback-giving behaviour.

Hypothesis 6: The personal characteristic conscientiousness is positively correlated with the frequency, quality of content and the consequential impact of the feedback process.

Openness to experience reflects the degree to which people enjoy rich, varied and novel experiences. Krasman was unable to establish a correlation between openness and active feedback-seeking behaviour but did find a correlation with passive feedback seeking. However, considering that feedback giving is an active behaviour and not directly linked to rich, varied and novel experiences, we did not hypothesize a relationship between openness to experience and elements of feedback giving behaviour.

Table 1 presents an overview of the hypotheses we tested.

	Feedback-giving behaviour		
	Frequency	Quality of content	Consequential impact
Positive self-efficacy	+	+	+
Positive task perception	+	+	+
High on neuroticism	+	-	+
High on extraversion	+		+
High on agreeableness	+	+	+
High on conscientiousness	+	+	+

Table 1 Hypothesized relationships of personality traits and elements of feedback-giving behaviour

Method

To test the hypotheses we conducted a cross-sectional survey by administering digital questionnaires to trainer-trainee couples. Trainers and trainees answered different questionnaires. The trainer questionnaire contained questions about six independent variables: self-efficacy, task perception, neuroticism, extraversion, agreeableness and conscientiousness, and the trainee questionnaire contained questions about the three dependent variables: frequency, quality of content and consequential impact of the feedback process.

Study context
During their specialist training GP trainees in the Netherlands spend a prolonged period of time working in a general practice where they are supervised by the same GP. They work in the practice four days a week, and one day per week attend a day

release programme at the university. Although they work independently most of the time, trainees can ask their trainer for help and advice and arrangements can be made for direct or video observation of a patient consultation followed by feedback.

Participants
Each of the eight University Medical Centres in the Netherlands offers a postgraduate training programme for general practice delivered by the local GP training institution. Within the framework of a faculty development programme, trainers from the institutions at Groningen, Utrecht and Rotterdam participated in our study. We asked first-year trainees and their trainers to fill in a digital questionnaire in the trainees' second or third month of training. During the first months of training trainees are usually observed fairly frequently, because trainers are eager to gauge their competence. We sent an invitation to participate in the study to 248 trainees (Rotterdam 86, Utrecht 98 and Groningen 64) who had started training in September 2011 or March 2012. 183 trainer-trainee couples signed for informed consent (Rotterdam 73, Utrecht 73 and Groningen 37). Two groups of trainer-trainee couples, one starting training in September 2011 and the other in March 2012, were asked to participate in the survey at the end of October or in November 2011 and at the end of April or in May 2012, respectively. Halfway these periods non-responders received a digital reminder. The study was approved by the ethics review board of the Dutch Association for Medical Education. All participants signed for informed consent.

Questionnaire
All four researchers (EP, AK, HM and CvdV) contributed to the development of the questionnaires. In July 2011 the questions were scrutinized by a group of experts/researchers consisting of two educational scientists, two GPs and an obs/gyn specialist. The feedback from the experts was used to adjust the questionnaires.

Dependent variables
Questions about trainers' feedback-giving behaviour in relation to the three dependent variables (frequency, quality of content and consequential impact) were put to trainees because we were more interested in trainees' perceptions of the feedback they received than in trainers' perceptions of the feedback they provided. Six questions about 'frequency' (Table 2) asked about the number of observations and the time spent on observations and feedback discussions. By multiplying the number of observations by the mean number of minutes per observation we obtained the total time (in minutes) spent on observation and feedback during the first two months of training. We adapted the measure of '*quality of content*' from the study by Adams[14] to the setting of our study and translated the questions into Dutch. This resulted in twelve questions with five-point Likert scales. Based on recommendations

from a review by Archer[23], we measured *'consequential impact'* using three questions about the linkage between feedback and trainees' learning goals, the possibility for trainees to reflect and linkage of new with earlier feedback, all with a five-point Likert scale (Table 2).

Measure	Sample item
Feedback-giving behaviour – frequency	1. How many times were you observed (live or via video) during a consultation you performed in the last two months? 2. How long were you observed on these occasions? Give the average time in minutes, excluding the time for the feedback discussion. 3. Give the average time in minutes for the feedback discussion. 4. How many times were you observed during a home visit or night shift in the last two months? 5. How long were you observed on these occasions? Give the average time in minutes, excluding the time for the feedback discussion. 6. Give the average time in minutes for the feedback discussion. (Answers in numbers [number of times or minutes])
Feedback-giving behaviour – quality of content	1. When I have performed poorly, my trainer provides detailed information about my performance. 2. When I have performed poorly, my trainer tells me specifically which aspects of my performance are in need of improvement. 3. When I have performed well, my trainer provides detailed information about my performance. 4. When I have performed well, my trainer tells me specifically which aspects of my performance were performed well. 5. I would like to receive more detailed feedback from my trainer. 6. After a discussion with my trainer I do not know exactly what I did well. 7. After a discussion with my trainer I do not know exactly what I did wrong. 8. My trainer tells me when I perform well. 9. My trainer gives me compliments. 10. My trainer tells me when I do not perform well. 11. My trainer expresses dissatisfaction when I do not perform well. 12. My trainer tells me when I have done something wrong. (Answers on 5-point Likert-scale: 1 = never; 5 = always)
Feedback-giving behaviour – consequential impact	1. The feedback I receive relates to my personal learning goals. 2. I get the opportunity to respond to feedback I receive. 3. The feedback I receive is linked to feedback I received on previous occasions. (Answers on 5-point Likert-scale: 1 = never; 5 = always)

Table 2 Items relating to the dependent variables

Independent variables

The trainer questionnaire contained questions about self-efficacy and task perception and questions from parts of a big-five questionnaire[21]. Task perception and self-efficacy were measured by three questions relating to preconditions for the feedback process (Table 3).

The questions about the personality traits were based on the Dutch version of the revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI). There were seven questions for each personality trait, which were mixed and anchored to either of the two extremes of a trait (Table 4).

Measure	Sample item
Self-efficacy	Indicate to what extent you consider yourself able to: 1. Arrange for observation of a patient consultation performed by your trainee. 2. Schedule time for a feedback discussion after an observation. 3. Ensure that the trainee translates feedback into learning goals, in other words ensure that your trainee actually uses the feedback you give to enhance his/her learning. (Answers on 5-point Likert-scale: 1= not at all; 5= totally)
Task perception	Indicate to what extent you consider the following activities to be a part of your tasks as a trainer: 1. Arrange for observation of a patient consultation performed by your trainee. 2. Schedule time for a feedback discussion after an observation. 3. Ensure that the trainee translates feedback into learning goals, in other words ensure that your trainee actually uses the feedback you give to enhance his/her learning. (Answers on 5-point Likert-scale: 1= not at all; 5= totally)

Table 3 Items for the independent variables self-efficacy and task perception

Analysis

For the continuous variables we calculated means and SDs. For self-efficacy, task perception, quality of content and consequential impact, the Likert type scores are presented as percentages after dichotomization (1-3 (fully) disagree or partly disagree versus 4 and 5 (fully) agree). The scores of negatively formulated items were recoded. Scores were aggregated per item by calculating the mean sum score for the original data, i.e. without dichotomization. To test for internal consistency we calculated Cronbach's alpha and conducted factor analysis for the Likert scores. The six questions about frequency of feedback were also put to the trainers, and trainers' and trainees' answers were analyzed using Pearson correlations. The trainers' answers were only used to check the quality of the data.

After bivariate analysis to test the hypotheses using Pearson correlation, we performed multiple regression analysis. An alpha level of .05 was considered to be significant.

Measure	Sample item
Neuroticism	1. I get upset easily 2. I tell myself that I am in trouble 3. I always see a ray of hope 4. I always fear for the worse 5. I quickly put aside setbacks 6. I can take a few blows 7. I quickly rack my brains over something
Extraversion	1. I make people smile 2. I keep in the background 3. I avoid company 4. I love big parties 5. I am afraid of new encounters 6. I start conversations 7. I prefer to be on my own
Agreeableness	1. I make efforts for other people 2. I take into account the interests of others 3. First of all, I think about myself 4. I take into account the feelings of others 5. I use others for my own goals 6. I impose my will on others 7. I respect other people's opinion
Conscientiousness	1. I do things without planning 2. I do things at the last minute 3. I make my work on time 4. I work in a fixed pattern 5. I want everyting exactly right 6. I leave my work unfinished 7. I am always well prepared
	All answers on 5-point Likert-scale: 1 = this totally applicable to me; 5 = this is totally not applicable to me

Table 4 Items for the independent variables neuroticism, extraversion, agreeableness and conscientiousness

Results

Descriptive statistics

Of the total of 248 trainer-trainee couples that were eligible for inclusion in the study, 183 gave informed consent and received questionnaires. Sixty-two couples (34%) completed the questionnaires (Groningen 43%, Utrecht 37%, Rotterdam 26%). The rather low response rate is attributable to the use of trainer-trainee couples as the unit of analysis which meant exclusion of a couple if data for one member was missing (Figure 1).
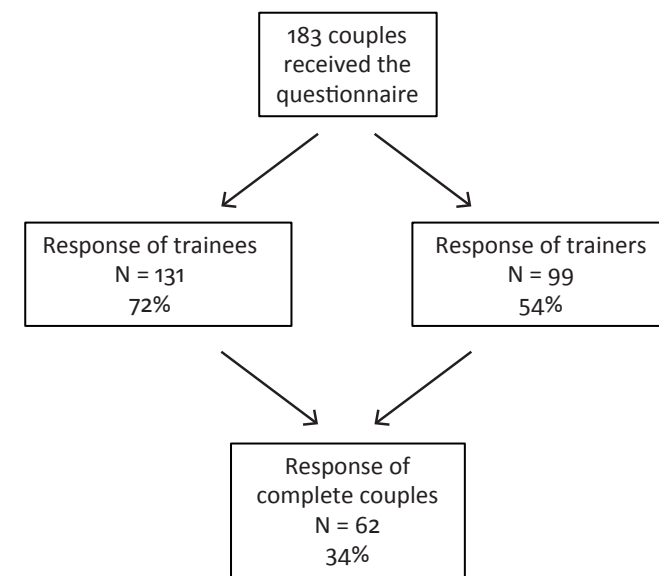
Figure 1 Flowchart of response rates

Dependent variables

Two researchers (EP and HM) examined extreme scores on the variable frequency, because due to misinterpretation of questions 2, 3, 5 and 6 (Table 2) some participants had multiplied the answers to questions 1 or 4 by the average number of minutes. When both researchers were certain this had happened, the mistake was corrected. The results show huge variation in the number of minutes of observation and feedback over two months. The maximum of 3090 minutes means in practice 1.6 hours of observation and feedback per day (based on 4 days a week for 8 weeks). There were significant correlations between the answers of trainees and their trainers.

'Quality of content' and 'consequential impact' were examined on a 5-point Likert scale. Table 5 shows scores of 3.57 and 3.94, respectively. The alpha for quality of content was high. Factor analysis was performed because of the low alpha for consequential impact. It showed that the items appeared to represent one construct. The last column in table 5 shows that only a minority of trainees gave high scores (score ≥ 4.00; average of 12 questions) on quality of the content of feedback, whereas almost 70% indicated that their trainers took steps to ensure the consequential impact of feedback (score ≥ 4.00; average of 3 questions).

	Mean	Minimum	Maximum	SD	Correlation with answers given by trainers	α	factor	(fully) agree (%)
Frequency	678*	70*	3090*	573*	.36**			
Quality of content	3.57	2.25	4.85	.44		.86		17.7
Consequential impact	3.94	2.67	4.67	.42		.47	1	69.4

* minutes over two months

**P < .01

Table 5 Descriptive statistics for the dependent variables

Independent variables

Table 6 shows the descriptive statistics for the independent variables. All these variables were examined on a 5-point Likert scale. The individual scores of trainers are the average of seven questions per personality trait (Table 4). Because of the low alpha for task perception we performed factor analysis, which showed that the questions represented one component. So, despite relatively low internal consistency, the items appeared to be related to one construct. The last column shows that at least three-quarters of the respondents did not agree that it was their task or felt confident to ensure preconditions for feedback (both score ≥ 4.00; average of 3 question), such as arranging for observation and feedback or supporting trainees in translating feedback into learning goals.

Correlation of dependent and independent variables

We found four significant correlations between dependent and independent variables (Table 7). Task perception and neuroticism showed positive correlations with frequency and quality of feedback.

	Mean	Minimum	Maximum	SD	α	Factor	(fully) agree (%)
Self-efficacy	3.43	2.33	5.00	.53	.71		19.4
Task perception	3.50	2.33	4.67	.49	.30	1	24.2
Neuroticism	2.22	1.00	3.29	.44	.78		
Extraversion	3.60	2.57	4.71	.53	.78		
Agreeableness	3.89	2.71	4.71	.35	.63		
Conscientiousness	3.52	2.57	4.57	.52	.81		

Table 6 Descriptive statistics for the independent variables

	Frequency	Quality of content	Consequential impact
Positive self-efficacy			
Positive task perception	.30*	.34**	
High on neuroticism	.33**	.31*	
High on extraversion			
High on agreeableness			
High on conscientiousness			

*P < .05

** P < .01

Table 7 Significant correlations between independent and dependent variables

To examine the potential interdependence of these correlations we performed multiple regression analysis. This showed that for frequency of feedback the independent variables task perception and neuroticism were correlated. As only the influence of task perception remained significant, the correlation between frequency and neuroticism depended on task perception. For quality of content things were the other way round with task perception depending on neuroticism.

Discussion

The purpose of this study was to contribute to the literature on feedback-giving behaviour by adding insights into factors within the person of the feedback giver. Task perception and the personality trait neuroticism were found to influence two elements of the feedback process: the frequency of feedback and the quality of feedback content. The results appear to support the conclusion that trainers who consider it to be their task to create favourable conditions for observation and feedback, are likely to show a higher frequency of feedback as well as better quality of feedback content. This is in line with hypothesis 2, but the results provide no evidence for the postulated effect on consequential impact. This is interesting in light of the finding that over 75% of the participating trainers disagreed with the statement that creating preconditions for observation and feedback was part of their task. This may offer a key to improving observation and feedback in general practice training. Earlier research showed that trainers who take an active attitude towards feedback giving are able to activate inactive trainees[15]. The present results, however, indicate that a positive task perception is prerequisite for trainers to observe and provide feedback more frequently.

Our results indicate a relationship between a neurotic personality of the trainer and frequency of observation, implying that emotionally stable trainers observe less frequently. This is in line with our hypothesis and with the literature on feedback seeking. Krasman demonstrated that neurotic persons seek more feedback, probably to alleviate a sense of insecurity[6]. Our results add to this that trainers with a more neurotic personality tend to observe more frequently. They may feel insecure leaving their patients in the care of trainees and are eager to ensure that their patients are safe. As a consequence the relative frequency of observation and feedback by emotionally stable trainers is lower. It should be noted that the correlation between neuroticism and the quality of the content of feedback turned out to run in opposite to the direction we had hypothesized. Our hypothesis stated that neurotic people are easily frustrated, irritable and prone to react violently[22] and that this detracts from the quality of feedback, whereas emotionally stable trainers provide better quality feedback. The findings, however, turned out to be the other way round. Apparently, feelings of insecurity had a positive effect on the quality of feedback while more stable personalities seemed more likely to leave matters to others. We found no evidence for an impact of neuroticism on consequential impact. The results support the conclusion that trainers with a more neurotic personality tend to give more feedback and that this feedback is of better quality.

Task perception and neuroticism were found to be interdependent. In other words, of the trainers with a high task perception a large group had a more neurotic personality or alternatively of the trainers with a more neurotic personality a large group had a high task perception. Feelings of insecurity might be involved too. Whereby, trainers with a more neurotic personality may be more inclined to comply with recommendations regarding observation from the training institution.

The results do not support effects of self-efficacy, extraversion, agreeableness or conscientiousness of trainers on elements of the observation and feedback process. This means that the results support neither our hypotheses nor results from the (feedback-*seeking*) literature[6,16]. This may be explained by the fact that feedback seeking and feedback giving are two different concepts relating to activities that are driven by different purposes. In feedback seeking the focus is on the person seeking feedback, who is intent on developing or demonstrating their own performance[8]. In feedback giving on the other hand the focus is on the recipient of the feedback, while the feedback is usually provided by someone in a professional capacity. The differences between our results and those of Kogan et al.[16] in relation to self-efficacy may be due to differences in operationalization. These differences warrant further research to clarify the potential effects of self-efficacy.

Recent research in the field of organizational psychology concluded that the feedback orientation of employees accounts for a substantial portion of the variance in the quality of coaching relationships between employees and their supervisors. In addition, empirical evidence supported a link between the coaching relationship and actual coaching behaviours, with perceptions of the coaching relationship accounting for significant variance in reports of actual coaching behaviour[24]. This shows that feedback recipients play a prominent role in the way feedback is given. Based on our findings and the literature on feedback seeking[8,10,11,25] we can conclude that feedback-seeking and feedback-giving behaviours constitute a highly complex phenomenon in which several actors and factors are interacting in complex and intricate patterns. Moreover, the process appears to be shaped by the hierarchical nature of the relationship between feedback seeker and feedback giver, whether this is a trainee and a trainer, an employee and a supervisor or otherwise.

The findings from this study are affected by limitations of the research method. Our failure to find any correlation between consequential impact and independent variables may be explained by the low alpha of consequential impact and the high scores on this variable from almost 70% of the participants, suggesting that a large majority of the trainees were convinced that their trainers took steps to further the consequential impact of feedback. These high scores, however, make it difficult to measure correlations with independent variables. Additionally, the high scores on consequential impact run counter to previous research showing that only 12% of trainee-trainer couples formulated an action plan based on feedback on observed performance[17]. A possible explanation may be that although no action plan was formally recorded, trainees nevertheless perceived that their trainers expected them to act upon feedback they had received. It is also important to notice the low alpha

for the explanatory variable task perception and the wide range of the scores on this variable while factor analysis showed that it was one component. A final limitation is the rather low response rate due to the requirement that answers had to be obtained from both members of trainer-trainee couples. Unfortunately, despite satisfactory response rates for trainees and trainers separately, many responses could not be analyzed because data for one member of a trainer-trainee couple was missing.

One of the strengths of this study is that participants were recruited from three different institutions for general practice, which strengthens the generalizability of the results to all programmes for postgraduate training in general practice in the Netherlands. Extrapolation of the results to other specialties or disciplines, such as organizational psychology, will require further research.

Our findings have practical implications as well. Because of its influence on feedback-giving behaviour, task perception should be a focus of training for trainers. Faculty development activities in postgraduate medical education often focus on (didactic) skills. The results suggest, however, that it may not be so much deficiencies in didactic skills but rather shortcomings in task perception that prevent trainers from engaging in observation and feedback.

This study expands on the literature on feedback giving by adding two factors that impact on feedback-giving behaviour: trainers' neuroticism and task perception.

References

1.   Shute, V.J. (2008). Focus on formative feedback. *Review of Educational Research, 78,* 153 – 189.
2.   Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77,* 81 – 112.
3.   Vancouver, J.B. & Morrison, E.W. (1995). Feedback inquiry – The effect of source attributes and individual differences. *Organizational Behavior and Human Decision Processes, 62,* 276 – 285.
4.   VandeWalle, D., Goutam, C.N., Shankar, G. & Brown, S.P. (2000). An integrated model of feedback-seeking behavior: Disposition, context, and cognition. *Journal of Applied Psychology, 85,* 996 – 1003.
5.   Madzar, S. (1995). Feedback seeking behaviour: A review of the literature and implications for HRD practitioners. *Human Resource Development Quarterly, 6,* 337 - 349.
6.   Krasman, J. (2010). The feedback-seeking personality: Big five and feedback-seeking behavior. *Journal of Leadership & Organizational Studies, 17,* 18 - 32.
7.   Ashford, S.J. & Cummings, L.L. (1983). Feedback as an individual resource – personality strategies of creating information. *Organizational Behavior and Human Performance, 32,* 370 – 398.
8.   Teunissen, P.W., Stapel, D.A., Vleuten, van der, C., Scherpbier, A., Boor, K. & Scheele, F. (2009). Who wants feedback? An investigation of the variables influencing residents' feedback-seeking behaviour in relation to night shifts. *Academic Medicine, 84,* 910 – 917.
9.   Sargeant, J., Mann, K., Sinclair, D., Vleuten, van der, C. & Metsemakers, J. (2007). Challenges in multisource feedback: intended and unintended outcomes. *Medical Education, 41,* 583 – 591.
10. Bok, G.J., Teunissen, P.W., Spruijt, A., Fokkema, J.P.I., Beukelen, van, P., Jaarsma, A.D.C. & Vleuten, van der, C.P.M. (2013). Clarifying students' feedback-seeking behaviour in clinical clerkships. *Medical Education, 47,* 282 – 291.
11. Crommelinck, M. & Anseel, F. (2013). Understanding and encouraging feedback-seeking behaviour: a literature review. *Medical Education, 47,* 232 – 241.
12. Pelgrim, E.A.M., Kramer, A.W.M., Mokkink, H.G.A. & Vleuten, van der, C.P.M. (2013). Reflection as a component of formative assessment appears to be instrumental in promoting the use of feedback; an observational study. *Accepted.*
13. Sargeant, J.M, Mann, K.V., Vleuten, van der, C.P. & Metsemakers, J.F. (2009). Reflection: a link between receiving and using assessment feedback. *Advances in Health Sciences Education, 14,* 399 – 410.
14. Adams, S.M. (2005). Positive affect and feedback-giving behaviour. *Journal of Managerial Psychology, 20,* 24 – 42.
15. Pelgrim, E.A.M., Kramer, A.W.M., Mokkink, H.G.A. & Vleuten, van der, C.P.M. (2012). The process of feedback in workplace-based assessment: organization, delivery, continuity. *Medical Education, 46,* 604 – 612.
16. Kogan, J.R., Conforti, L.N., Bernagbeo, E.C., Durning, S.J., Hauer, K.E. & Holmboe, E.S. (2012). Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Medical Education, 46,* 201 – 215.
17. Pelgrim, E.A.M., Kramer, A.W.M., Mokkink, H.G.A. & Vleuten, van der, C.P.M. (2012). Quality of written narrative feedback and reflection in a modified mini-clinical evaluation exercise: an observational study. *BMC Medical Education, 12,* 97.
18. Bandura, A. (1977). *Social learning theory.* Englewood Cliff: Prentice-Hall.
19. Boendemaker, P.M.C., Schuling, M.H., Meyboom - de Jong, B., Zierstra, R.P. & Metz, J.C.M. (2003). Core characteristics of the competent general practice trainer, a Delphi Study. *Advances in Health Sciences Education, 8,* 111 – 116.
20. Fluit, C.R.M.G., Bolhuis, S., Grol, R., Laan, R. & Wensing, M. (2010). Assessing the quality of clinical teachers: a systematic review of content and quality of questionnaires for assessing clinical teachers. *Journal of General Internal Medicine, 25,* 1337 – 1345.
21. Costa, P.M. & McCrae, R.R. (1992). *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI) professional manual.* Psychological Assessment Resources: Lutz.
22. Mattheyws, G. & Deary, I.J. (1998). *Personality traits.* Cambridge University Press: Cambridge.
23. Archer, J.C. (2010). State of the science in health professional education: effective feedback. *Medical Education, 44,* 101 – 108.

24. Gregory, J.B. & Levy, P.E. (2012). Employee feedback orientation: Implications for effective coaching relationships. *Coaching: An International Journal of Theory, Research and Practice, 5,* 86 – 99.

25. Pelgrim, E.A.M & Kramer, A.W.M. (2013). How can medical education benefit from the evidence on learners seeking and using feedback? *Medical Education, 47,* 225 – 227.

Chapter 7

General discussion

General discussion

The studies that together constitute this thesis were undertaken to shed light on the workings of observation and feedback in workplace-based assessment. Our interest in this subject was triggered by reports of infrequent occurrence of observation of single patient encounters which we found in the international literature and more specifically in a recent evaluation of postgraduate training in general practice in the Netherlands[1-3]. The subject merits extensive study because of changes of curricular structure in workplace-based programmes and evidence of the importance of observation for workplace-based assessment and learning[4]. We specifically investigated: 1) the role of assessment instruments and 2) the role of the trainee and 3) the role of the trainer in the observation and feedback process. In this chapter we return to these three questions, weighing the strengths and limitations of the studies and proposing directions for future research. We also discuss practical implications of the findings for workplace-based learning and assessment.

Findings relating to the three main questions

The study of the process of observation and feedback in workplace-based assessment revealed three crucial steps (Chapter 3), summarized in a three-step model (Figure 1). Firstly, observation needs to be *organized*, because it is facilitated when it is embedded in practical routines. Secondly, the *quality* of feedback is important. In other words, the content and delivery of feedback should be suitable to initiate a learning process. The last step (*consequential impact*) is concerned with activities to ensure continuation of the learning process. These three steps together can optimize workplace-based assessment. Factors that influence the model will be discussed. To answer the three main questions of this thesis we describe how assessment tools, the trainee and the trainer impact the steps of the model. Anticipating on the final conclusions we can say that the model will be shown to be too simplistic.

Figure 1 Steps in the process of obtaining useful feedback through deliberate observation in the workplace

The assessment instrument

In medical education, a great deal of effort and time has been expended on developing reliable and valid instruments for workplace-based assessment, in particular for observation of single patient encounters[5,6]. The literature review in Chapter 2 shows that such instruments generally meet with a positive reception. Interestingly, despite the absence of clear criteria for feasibility, feasibility was the aspect of the instruments that received the most positive evaluations, and most formats were considered feasible. All the instruments included in the review had been adapted for a specific purpose, some for a particular medical specialty (rheumatology), some for a specific competency (professionalism), some for a specific task (palliative care), and this has probably enhanced their feasibility. None of the instruments resulted in demonstrable improvement of clinical skills or the quality of patient care. It seems safe to conclude that assessment instruments can have a supportive role in workplace-based assessment. Feasibility rates and the availability of many different instruments for single-encounter observation suggest that it is preferable to use an instrument that is tailored to the situation of interest. When an instrument is implemented, it is important that it should be evaluated.

The main question we sought to answer was: What is the role of an assessment tool in the feedback process based on observation of single patient encounters? It was established that an assessment tool can have a supportive function, but can it have other functions as well? The study in Chapter 3 supports the conclusion that an assessment instrument is *not a key factor* in optimizing learning from observation-based assessment. A comparison of GP trainees from two training institutions using different methods for observation of and feedback on trainee performance in patient encounters revealed no striking differences in the use of feedback.

The study described in Chapter 4 shows that trainer and trainee had a stronger impact on the use that was made of an assessment instrument than the instrument itself. The instrument provided space for three types of narrative comments, 1) reflection 2) feedback and 3) action plans based on the feedback. Feedback was entered frequently, reflections only in about half of the completed forms and action plans were very scarce. Although the layout of the instrument indicated where which type of comment was to be entered, users did not follow these instructions. Some used the whole instrument, while others systematically used the space for feedback only. Apparently, it was not the instrument but the trainer and the trainee that determined the feedback process.

A similar conclusion was reported by Crossley and Jolly, who criticized the use of instruments to shape the behaviour of feedback-givers without taking into account their expertise and capacities[7]. They warned that instruments might limit opportunities for observation and feedback and concluded that in workplace based assessment we should make the best possible use of the expertise of observers instead of forcing them into the straitjacket of the faculty's assessment programme, for instance by making it mandatory to use standardized instruments and response formats.

A recent study with a control group showed that written or verbal feedback facilitated by a feedback instrument resulted in improvement which was deemed attributable to the combination of direct observation and feedback[8]. However, the researchers did not discuss the instrument and merely concluded that weekly direct observation and feedback improved certain aspects of patient care.

In summary, we think the findings of our studies together with recent evidence from the literature support the conclusion that while an assessment instrument can be supportive it is probably not a decisive factor in obtaining useful feedback. It is the users of the instruments who are in control.

The trainee

Chapter 3 shows that trainees' responses to observation are primarily emotional. Trainees were apprehensive, indicated they behaved differently under observation and generally did not enjoy the experience. Although these emotional responses were shared by almost all trainees, trainees showed two distinct coping patterns. Some trainees engaged actively in feedback seeking, acknowledging, despite their misgivings, that feedback could be beneficial and help them improve their performance. Other trainees allowed their fears to prevail and shrank back from making arrangements for observation and feedback.

These behavioural patterns are best investigated by research in the area of feedback seeking. Although this type of research has quite a long history in organizational and social psychology[9], it has only recently come into its own in medical education[10]. Feedback seeking research sees the trainee as an active partaker in the feedback process who can initiate the first step of the three-step model (Figure 1). A recent review in organizational psychology concluded that feedback-seeking behaviour was affected by many personal, interpersonal and contextual factors[11]. In a commentary on that review[11] and a research paper[12] we argued that feedback-seeking behaviour is a highly complex phenomenon, with several actors and factors interacting in complex and intricate patterns[13].

The studies we conducted for this thesis demonstrate that the trainee is an active participant in the feedback process. In Chapter 5 we described a more detailed investigation of the influence of the trainee on observation and feedback. Differences between trainees appeared to affect the extent to which trainees learned from workplace-based assessment. Trainees who engaged more often in reflection on their performance seemed to be more likely to make an action plan based on feedback. This relates to the last step of the three-step model (Figure 1), which pertains to the learning effect of feedback, i.e. how the trainee uses feedback for learning.

Sargeant et al.[14] reported that reflection was stimulated by feedback that was inconsistent with the recipient's self-perceptions and that reflection was the process through which feedback was assimilated and therefore integral to decisions about acceptance and usage of feedback. Facilitated reflection upon feedback was identified as a positive factor for assimilation and acceptance. Although Sargeant's conclusions appear to confirm our results, it is important to note that unlike Sargeant, who investigated reflection on feedback received, we investigated trainees' reflections on their own performance[14]. This type of reflection is akin to self-assessment, not in the sense of judging oneself, but as a mechanism for on-going monitoring, identifying and redressing gaps in competence[15]. Both forms of reflection, however, seem to be instrumental in the assimilation of feedback (Chapter 5)[14].

The main outcome concerning the role of the trainee appears to be that trainees can take an active attitude towards observational assessment. Not only can they be active in the first step of the three-step model (Figure 1) by organizing opportunities for observation, they can also be active in the third step by reflecting on their performance (self-assessment), thereby increasing the likelihood that feedback is actually used for developmental purposes. In brief, the trainee can influence the first and third step of the model.

The trainer
The trainer is the other leading actor in the feedback process. Several studies in this thesis show that the trainer is crucial for learning through observation and feedback in the workplace. The feedback-seeking literature shows that trainers impact trainees' feedback-seeking behaviours, providing evidence for an effect of the quality of the trainer-trainee relationship as well as the trainer's leadership style, mood and expertise[11].

Chapter 3 shows that the avoidance behaviour of trainees who are aversive to observation and feedback can be counteracted by an active GP trainer. We identified two groups of trainers: those who take initiatives for observation and feedback and those who refrain from it. Trainers who actively promote observation and feedback will succeed most of the time, even if trainees are not cooperative. (Table 1).

The feedback-seeking literature shows that the personality of the trainer and factors of the trainer-trainee relationship influence trainees' feedback-seeking behaviour[11]. Our results show that the trainer can *overrule* the trainee's preferred behaviour in respect of observation. In other words, in the final instance it is the trainer who determines what happens, probably due to the hierarchical relationship between trainer and trainee.

The influence of the trainer on observational assessment in the workplace is highlighted in several chapters. Chapter 5 shows that only if the trainer is capable of providing specific written feedback, is there a chance that the trainee will provide

	GP trainer active in feedback giving	GP trainer inactive in feedback giving
Trainee active in feedback seeking	A great deal of activity in observation and feedback.	The GP trainer does not actively promote feedback, the trainee has to ask for it. Feedback will occur occasionally.
Trainee inactive in feedback-seeking	The trainee is reluctant to engage in observation, but the GP trainer makes it happen.	No observation or feedback without external pressure (from the training institution)

Table 1 Types of combinations of GP trainee and GP trainer

written reflections on performance. Apparently, trainees are not inclined to write down their thoughts unless the trainer sets an example. In this process the trainer takes the lead and acts as a role model.

The investigation of trainer characteristics (Chapter 6) shows an effect of the personality trait 'neuroticism' on observation and feedback-giving. Trainers with a more neurotic personality observe trainees more often and give better feedback, perhaps because they feel uncertain about entrusting patients to the care of a trainee and eager to ensure patient safety. Another important finding was the impact of trainers' task interpretation. It appears that trainers who see it as their task to create favourable conditions for observation and feedback (make arrangements for feedback and encourage trainees to use it) give more and better feedback. This corresponds with Korthagen's notions[16,17] about the professional development of teachers. His onion model (Figure 2) shows that professional development is not a passive process but that teachers are actively involved in decisions about the direction and process of their own teaching. The model shows that the environment, behaviour, competencies, beliefs, identity and mission of the trainer all play their part in the practice of trainers.

Our finding that trainers' interpretation of their teaching tasks influences their performance in observing and giving feedback corresponds with the 'professional identity' layer of the onion model, pertaining to questions like: 'Who am I as a trainer?', 'What is my professional role?' 'What kind of trainer do I strive to be?'. Korthagen emphasizes the importance of these 'inner layers' of the onion model. The importance of the trainer as a person is overlooked only too often in professional development.
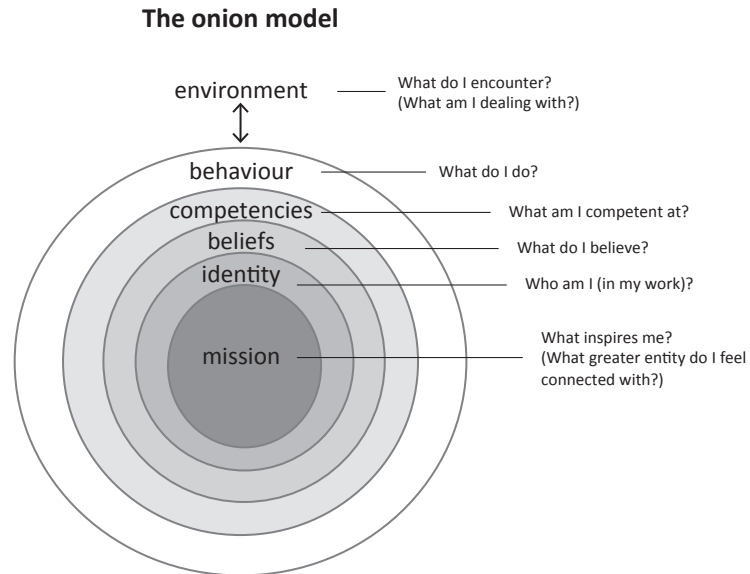
The onion model

environment — What do I encounter?
(What am I dealing with?)

behaviour — What do I do?

competencies — What am I competent at?

beliefs — What do I believe?

identity — Who am I (in my work)?

mission — What inspires me?
(What greater entity do I feel
connected with?)

Figure 2 Korthagen's onion model; the importance of the mission, identity, beliefs, competencies and behaviours of trainers for their teaching

Conclusion

To optimize workplace-based learning and assessment by the use of observation and feedback, trainers and trainees should proceed through the three-step model (Figure 1). It is not the assessment instrument that is important but the way the assessment is conducted by the key participants. Whilst trainees and trainers both impact the process, the trainer is of crucial importance in activating and motivating trainees to receive feedback and use it to promote learning. If the trainee is reluctant to play an active role, the trainer can activate the trainee, but unfortunately these roles are not easily reversed. It is not self-evident that an active trainee is able to activate an inactive trainer, probably due to the hierarchical nature of the trainer trainee relationship. It should be noted that although trainers are capable of optimizing the feedback process, they are only likely to do so if they accept this as a component of their teaching task.

Figure 3 shows the elements influencing the three-step model. Apart from the factors described in this thesis and shown in Figure 3, the literature also refers to 'climate' (the atmosphere in which feedback is given)[18] and 'culture'[19,20] as important elements. When an appropriate time and location are arranged for feedback, the atmosphere is relaxed and respectful and there is a negotiated agenda, the feedback process will probably be more effective[18].
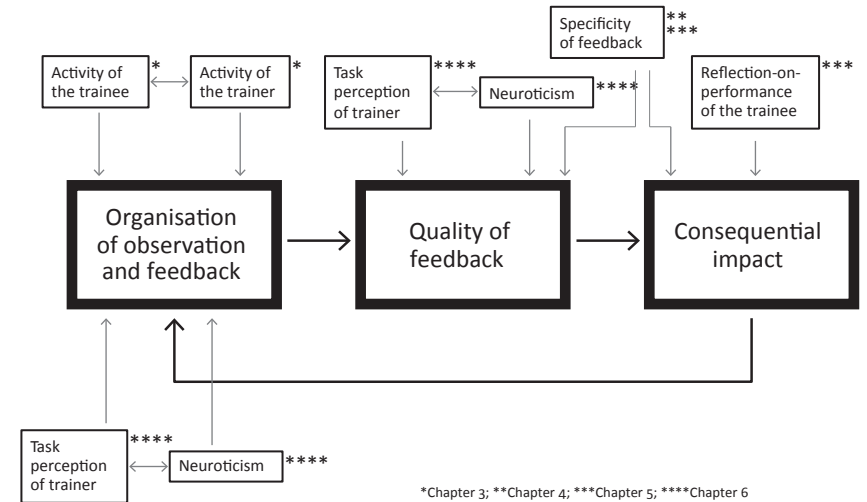

*Chapter 3; **Chapter 4; ***Chapter 5; ****Chapter 6

Figure 3 The elements we found of influence on the three-step-model

Culture refers to 'unwritten rules' and expectations of doctors (or trainees). Sargeant[20] showed that the culture in which doctors perform is one that values 'reputation' and 'being positively viewed by colleagues'. According to Archer, this medical culture needs to develop into a 'feedback culture' in which feedback is embedded implicitly and explicitly in all activities and in which trainees give feedback to trainers and trainers give feedback to trainees[19]. Early training and experience, such as peer feedback, may help to accomplish this desired cultural change over time. Archer also posited that feedback opportunities must be actively sought and encouraged.

Observation of patient consultations and programmatic assessment

In the introduction of this thesis we described Van der Vleuten et al.'s model for programmatic assessment[21]. The purpose of this model is to maximally support 'on-going learning' while at the same time enabling robust decisions about progress. Training and assessment activities should be arranged so as to maximally support on-going learning. Observation of patient consultations fits perfectly within this model[21]. The investigations we conducted of factors influencing the process of observation and feedback showed that this process is complex and influenced by numerous factors some of which can and some of which cannot be controlled.

Considering that our studies addressed only a limited part of the model, it is only to be expected that the total model will be subject to the effects of at least as many factors. After examining an assessment programme based on the model of programmatic assessment, Driessen et al.[22] concluded that assessment programmes should provide feedback, be simple in structure and be learner led, while users should be thoroughly familiarized with the programme. Our conclusion with regard to observational assessment is quite similar: feedback is essential, the feedback process should be user friendly and feedback providers and recipients should be familiarized with it. This can be achieved through training, but training in how to conduct observational assessment is not enough, the inner circles of the onion model[16,17] needs to be addressed as well and the programme should fit with the beliefs and identities of the users. We partly support Driessen et al.'s[22] conclusion that a programme of assessment should be learner led, for example by allowing learners to select areas for observation and feedback, but, based on the study in Chapter 3, we would add that active interference by the trainer is essential to counteract inactivity of trainees and capture their blind spots[15].

Methodological considerations

This thesis integrates several studies which all contribute to the central objective: How to optimize observation and feedback in workplace-based assessment in order to enhance learning? In Chapters 3-6 we investigated the influence of several key elements and actors in the natural environment of postgraduate training. This approach has the merit of providing evidence from the real world rather than from a controlled environment, thereby enhancing the relevance of the findings[23]. The strengths and limitations of the studies are associated with this real world perspective. The main strength of the research is its relevance to the practice of medical education. Education and training of future medical specialists is crucial for maintaining high standards of healthcare. Postgraduate training takes place mainly in the workplace, where learning depends on observation, feedback and reflection. This applies not only to postgraduate education but also to undergraduate clinical training and to continuing development of practising professionals. The profession is developing, so professionals must learn and continue to develop with it.

Apart from practical relevance, the research has theoretical relevance. Currently, workplace-based learning and assessment feature prominently in the international literature[2,7,21,22,24,25]. While it is important to investigate whether assessment works, it is at least as important to clarify *how* it works and to study theoretical frameworks, like the model of programmatic assessment[21]. This thesis builds on this theory by providing evidence regarding factors influencing the complex field of observation

and feedback in workplace-based learning and assessment thereby adding substance to the model of programmatic assessment.

Another strength of the research is its methodological rigour. This is evident from the use of different methodologies and types of data. The mixture of quantitative and qualitative research allowed for cross-comparison of results and permitted triangulation of findings.

Some limitations of the research should be considered as well. Because the studies were conducted within postgraduate training in general practice in the Netherlands, the generalizability of the empirical results seems not entirely certain. Although the studies deal with generic themes in medicine, such as observation of patient consultations and feedback, the specifics are associated with one particular setting. This warrants caution in claiming that the findings will hold for trainers and trainees in other specialties, undergraduate students and their supervisors, or medical education outside the Netherlands. Further limitations are inherent in the nature of qualitative research. Although the interview study was based on theory and the data were interpreted by multiple researchers, bias cannot be ruled out. The use of questionnaires can be a source of bias as well. Although multiple researchers were involved in the study and we invited independent colleagues to review the questionnaire, the instrument may have elicited socially desirable answers. Furthermore, the questionnaire examined participants' views not their actual performance or personality. A final potential source of bias is the self-selection of participants. The fact that participants had to choose to participate and (logically) could not be forced to do so does contribute to the power of the study.

Recommendations for future research

The research presented in this thesis was aimed at clarifying observation and feedback in workplace-based assessment. The results have been translated into a simplistic model (Figure 1) of the feedback process and an extensive model (Figure 3) of factors influencing this process. The results contribute to a better understanding of observational assessment in workplace-based learning, and also open up avenues for further research. Future research should focus on the trainer, the trainee, the interaction between them and the environment around them.

Firstly, we concluded that the trainer is a powerful influence in the process of observation and feedback and in Chapter 6 we built on the existing knowledge about feedback by showing that factors within the personality of the trainer influence feedback-giving behaviour. Nevertheless a great deal remains to be learned about the feedback-giver. Our results are a first step, which should be followed up by further research into factors influencing feedback-giving behaviour. It would be interesting to

better understand what influences feedback-giving behaviour. The onion-model of Korthagen (Figure 2)[17] gives an overview of areas that influence the behaviours of trainers. Future research should investigate how these different layers of the onion-model influence feedback-giving behaviour. Which trainers do give feedback and which trainers do not and why so? Is the environment decisive? Or are they not competent? Or is feedback giving not something that they believe in? When we understand these mechanisms we are able to change for the better. Firstly we can train feedback-givers in a more effective way. Secondly, we then are able to modify assessments instruments in a way that better fits for the users. If we know more about the feedback-giver we can shape the assessment instruments in such a way that it can support them. In contrast to forcing the feedback-giver in the format of the instrument[7].

The influence of the trainee and the relationship between trainer and trainee also appear to be an area that is worth investigating further. Alignment with feed-back-seeking research could be useful here. It would be interesting to investigate which trainer fits best with which trainee in the context of feedback-giving and feed-back-receiving behaviour. In this context also the feedback climate and feedback culture are interesting to investigate further. What affects a positive feedback climate or feedback culture? What is the influence of the feedback-giver and the feedback-receiver in this? And what is the influence of the faculty?

We conclude by stressing the importance of familiarizing the users of workplace-based assessment with this type of assessment. We need to determine how this can best be done. What kind of training is needed? Do we need to train the trainers only, or should the trainees be trained as well? What are the long-term effects of training? And how do we address beliefs and identities?

Our studies were conducted in the setting of general practice in the Netherlands. Further research should determine the applicability of the findings to other medical specialties, undergraduate medical education and perhaps even other workplace-based training settings outside the medical domain. It is important to keep in mind that in the setting of our studies one trainer and one trainee work together for a whole year, while in other settings trainees have multiple supervisors and receive feedback from a variety of sources. Probably the medical educational domain could learn from other domains and vice versa. For feedback-seeking behaviour this conjunction is already found between medical education and organizational psychology. Perhaps there are more commonalities between these domains and we can learn from each other's expertise in multiple areas. Another exchange of knowledge could be the education in Universities of Applied Science in the Netherlands. Almost all these institutions use a competence based curriculum with a large share of workplace-based learning. Do they recognise our results and clarifications about observation and assessment feedback? Finally, another context

in which observation, feedback and reflection is of frequent occurrence is in sports. What can we learn from athletes and their coaches who constantly observe, watch videos, talk, give and get feedback and reflect about their training and matches? And what can they learn from our evidence?

With respect to research methods, we could learn from anthropological research methods. Participating and observing what happens in practice could give further insight in what is actually happening in the workplace. Also conversation analysis could give us new insights. We now only investigated what people said what happened, or data which were listed on the assessment instrument. This gives us only a limited view of reality. Anthropological research methods can complement our data to broaden the image we have.

Practical recommendations

The findings of this thesis have practical implications. An important insight is the relative unimportance of assessment instruments. They do have a supportive function but they are not at the heart of the feedback process. It would seem more worthwhile to devote energy and time to the feedback-giver and the feedback-recipient, i.e. the trainer and the trainee. It should be noted that the influence of the trainer is greater than that of the trainee. Trainer and trainee need to be trained for and familiarized with feedback and observation. And attention must be paid to 'the trainer as a person', including his/her beliefs and identity. Although this is often overlooked in professional development, it is of paramount importance.

Moreover, it is important to carefully determine and evaluate which trainer and trainee are linked to each other. As we have seen in chapter 3 and Table 1 of this chapter, couples could influence each other in a positive or in a negative way. For the faculty it is important to respond on this.

The results of our thesis do not provide a clear-cut definitive answer to the question of how observation and workplace-based assessment are best conducted. What is clear is that the process should be tailored to each single workplace-based setting. In designing an assessment programme, it is vital to place observation at the centre[18,20]. Account should also be taken of the fact that feedback is better accepted when the feedback-recipient is directly observed by the feedback-giver. Also, the users of the assessment programme need to be trained. They are the key players and the success or failure of the programme depends on them. After an assessment programme is developed, ample attention must be paid to a careful implementation and a thorough evaluation process. Only when these conditions are met can work-place-based assessment be made effective.

In conclusion

In our endeavours to bring clarity to observation and feedback in workplace-based assessment, we examined the role of assessment tools, trainees and trainers. An important finding is that the human actors are the heart of the feedback process, with the trainer exerting even more influence than the trainee. Assessment instruments are of minor importance. The process of observation and feedback proved to be an important but highly complex phenomenon in which several actors and factors interact in complex and intricate patterns which determine whether observation takes place and if it does whether feedback is appropriate and put to effective use for further learning and development.

References

1. Fromme, H.B., Karani, R. & Downing, S.M. (2009). Direct observation in medical education: a review of the literature and evidence for validity. *Mount Sinai Journal of Medicine, 76,* 365 – 371.
2. Norcini, J.B. & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher, 29,* 17.
3. Velden, van der, L.F.J. & Batenburg, R.S. (2011) De opleiding tot huisarts opnieuw beoordeeld. Een onderzoek onder huisartsen in opleiding en alumni. http://www.nivel.nl/pdf/rapport-kwaliteit-huisart-senopleiding.pdf. [retrieved: February 2013].
4. Sargeant, J., Mann, K. & Ferrier, S. (2005). Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Medical Education, 39,* 497 – 504.
5. Pelgrim, E.A., Kramer, A.W., Mokkink, H.G., Elsen, van den, L., Grol, R.P. & Vleuten, van der, C.P. (2011). In-training assessment using direct observation of single-patient encounters: a literature review. *Advance in Health Sciences Education, 16,* 131 – 142.
6. Kogan, J.R., Holmboe, E.S. & Hauer, K.E. (2009). Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA, 302,* 1316 – 1326.
7. Crossley, J & Jolly, B. (2012). Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. *Medical Education, 46,* 28 – 37.
8. Shelesky, G., D'Amico, F., Marfatia, R., Munshi, A. & Wilson, S.A. (2012). Does weekly direct observation and formal feedback improve intern patient care skills development. A randomized controlled trial. *Family Medicine, 44,* 486 – 492.
9. Ashford, S.J. & Cummings, L.L. (1983). Feedback as an individual resource – personal strategies of creating information. *Organizational Behavior and Human Performance, 32,* 370 – 398.
10. Teunissen, P.W., Stapel, D.A., Vleuten, van der, C., Scherpbier, A., Boor, K. & Scheele, F. (2009). Who wants feedback? An investigation of the variables influencing residents' feedback-seeking behavior in relation to night shifts. *Academic Medicine, 84,* 910 – 917.
11. Crommelinck, M. & Anseel, F. (2013). Understanding and encouraging feedback-seeking behaviour: a literature review. *Medical Education, 47,* 232 – 241.
12. Bok, G.J., Teunissen, P.W., Spruijt, A., Fokkema, J.P.I., Jaarsma, A.D.C. & Vleuten, van der, C.P.M. (2013). Clarifying students' feedback-seeking behaviour in clinical clerkships. *Medical Education, 47,* 282 – 291.
13. Pelgrim, E.A.M. & Kramer, A.W.M. (2013). How can medical education benefit from the evidence around learners seeking and using feedback. *Medical Education, 47,* 225 – 227.
14. Sargeant, J.M., Mann, K.V., Vleuten, van der, C.P. & Metsemakers, J.F. (2009). Reflection: a link between receiving and using assessment feedback. *Advances in Health Sciences Education, 14,* 399 – 410.
15. Eva, K.W. & Regehr, G. (2005). Self-assessment in the health professions: a reformulation and research agenda. *Academic Medicine, 80,* S46 – 54.
16. Korthagen, F.A.J. (2009). Professional learning from within. *Studying Teacher Education, 5,* 195 – 199.
17. Korthagen, F.A.J. (2004). In search of the essence of a good teacher: towards a more holistic approach in teacher education. *Teaching and Teacher Education, 20,* 77 – 97.
18. Hewson, M.G. & Little, M.L. (1998). Giving feedback in medical education: verification of recommended techniques. *Journal of General Internal Medicine, 13,* 111 -116.
19. Archer, J.C. (2010). State of the science in health professional education: effective feedback. *Medical Education, 44,* 101 – 108.
20. Sargeant, J., Mann, K., Sinclair, D., Vleuten, van der , C. & Metsemakers, J. (2008). Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Advances in Health Sciences Education, 13,* 275 – 288.
21. Vleuten, van der , C.P.M., Schuwirth, L.W.T., Driessen, E.W., Dijkstra, J., Tigelaar, D., Baartman, L.K.J. & Tartwijk, van, J. (2012). *Medical Teacher, 34,* 205 – 214.
22. Driessen, E.W., Tartwijk, van, J., Govaerts, M., Teunissen, P. & Vleuten, van der, C.P. (2012). The use of programmatic assessment in the clinical workplace: a Maastricht case report. *Medical Teacher, 34,* 226 – 231.

23. Buckley, G. (1998). Partial truths – research papers in medical education. *Medical Education, 32,* 1-2.
24. Govaerts, M.J.B., Schuwirth, L.W.T., Vleuten, van der, C.P.M. & Muijtjens, A.M.M. (2011). Workplace-based assessment: effects of rater expertise. *Advances in Health Sciences Education, 16,* 151 – 165.
25. McGill, D.A., Vleuten, van der, C.P. & Clarke, M.J. (2011). Supervisor assessment of clinical and professional competence of medical trainees: a reliability study using workplace data and a focused analytical literature review. *Advances in Health Sciences Education, 16,* 405 – 425.

Chapter 8

Summary

Samenvatting

Dankwoord

Curriculum Vitae

Summary

Clarifying observation and assessment feedback in workplace-based learning

In this thesis we examine factors influencing observation and feedback in workplace-based assessment and learning. The research was conducted in the setting of postgraduate training in general practice in the Netherlands between 2009 and 2012. After a general introduction in which we pay attention to theoretical aspects of workplace-based learning and assessment, five chapters present the results of studies on elements of observation and feedback in workplace-based learning and the findings are synthesized and discussed in the final chapter.

Chapter 1 introduces the starting point of this thesis: an educational evaluation among general practice trainees in the Netherlands showing that numerous trainees are rarely if at all observed during single patient encounters. Similar findings have been reported in the literature on undergraduate and postgraduate medical education. The introductory chapter starts by painting a broad picture of workplace-based assessment and underlying theoretical perspectives, in particular Miller's pyramid, formative and summative assessment, and the model of programmatic assessment of Van der Vleuten. Next the focus narrows to observation and feedback, and three elements are identified that are central to workplace-based learning and assessment: the *trainee,* who needs to develop; the *trainer,* who guides and assesses the learning process; and a real patient setting which cannot (and should not) be fully structured or standardized for assessment purposes but ensure fair assessment and useful feedback, probably using an *assessment tool*. The general objective of this thesis is to clarify observation and feedback in workplace-based assessment. To unravel this process we focus on assessment instruments, the influence of the trainee and the influence of the trainer.

Chapter 2 presents a review of the literature on instruments for assessing single patient-trainee encounters in the workplace aimed at helping trainees gain insight into the strengths and weaknesses of their clinical performance. Most of the instruments reviewed were developed for formative assessment. Generally, they were given a positive reception, their feasibility was deemed to be good and acceptable and reliability was achievable with ten encounters. Of many instruments the validity was not investigated, but the validity of the mini-CEX and the clinical evaluation exercise was supported by strong and significant correlations with other valid assessment instruments. Evidence for educational effects was not very convincing. None of the studies included in the review examined effects of instruments on learning, clinical

skills or the quality of patient care. There was substantial variation between instruments in the competencies assessed, rating scales, frame of reference and user training. Assessment characteristics remained implicit and interpretation was largely left to the trainers. The effective implementation of instruments for workplace-based assessment probably hinges on appropriate training for those using them.

Chapter 3 focuses on formal feedback to trainees after an observed consultation. The assumption underlying the use of feedback is that it makes trainees aware of shortcomings and motivates them to improve their performance. Trainees were interviewed about external conditions affecting feedback after observed encounters in workplace-based assessment. The results revealed several conditions for enhancing the effects of feedback: 1) observation and feedback are planned by the trainer and the trainee; 2) the content and delivery of the feedback are adequate; and 3) the trainee uses the feedback to set learning goals to guide learning. All trainees showed primarily emotional responses to observation, notably apprehension, but differed in the strategies they used to deal with it. While some avoided observation, others overcame their apprehension and actively sought observation and feedback. Avoidant trainee behaviour could be counteracted by an active trainer. Trainers also fell into two groups: those actively initiating and those refraining from observation and feedback.

The study in **Chapter 4** investigates written feedback after observed patient encounters entered in a form that was designed to meet the following general notions about feedback. Feedback is preferably narrative and specific, explicating where more work needs to be done. Feedback can be made more effective when recipients receive guidance on turning feedback into steps to improve performance. Feedback is preferably a two-way process with trainers providing comments and at the same time encouraging trainees to self-reflect on their performance. We studied the effects of an assessment instrument designed to stimulate trainers to provide written feedback, trainees to write self-reflections and trainer and trainee to write an action plan derived from the feedback and reflections. Trainers provided feedback more frequently than trainees wrote self-reflections, while action plans were very rare. The written comments were generally specific. Because in the context of this study a trainer and a trainee team up for the duration of one year and because trainer-trainee teams showed huge differences in the number of forms handed in, we examined the specificity of comments using trainer-trainee teams as the unit of analysis. Differences proved to be large, with some teams consistently writing specific comments, while others consistently failed to do so. Although the layout of the form indicated which comments were to be entered in which spaces, in the end it was up to the users how to use the form. Based on the substantial differences between trainer-trainee teams

we conclude that it are the users rather than the instrument that determines the feedback process. That is why it seems advisable to direct interventions to improve the educational effects of feedback not at feedback instruments but at the users of those instruments.

Chapter 5 builds on the results of Chapter 4 by examining the relationship between reflection, feedback and the formulation of an action plan. Despite the absence of empirical evidence, the literature suggests that (self-)reflection can have a positive effect on the use of feedback. We hypothesized that feedback and reflection present a cumulative sequence and that reflection impacts on the use of feedback. In other words, trainers who provide specific feedback are more likely to pay attention to trainees' reflections, and reflections in turn affect the use of feedback, with trainees who reflect on their performance being more likely to make use of the feedback they receive. We did this by examining the assessment forms used in Chapter 4. The results support both hypotheses. We found a cumulative sequence of feedback and reflection as well as an effect of reflection on the use of feedback. Specific reflections were found in combination with specific feedback only and they occurred more often in combination with specific action plans. This provides new evidence that (self-)reflection can promote the use of feedback. Apparently, a reflective attitude towards one's own performance plays an important role in applying feedback to enhance learning and improve performance. The feedback-reflection-action sequence might be used for faculty development in the area of feedback. Feedback has received considerable attention in (medical) education for quite some time, and trainers have learned a great deal about the importance of feedback and how it is best provided. The realization of the importance of reflection and its relationship with the use of feedback to set learning goals is a later development, however, and these notions have not yet found their way into the standard feedback procedures of many trainers.

In **Chapter 6** we examine the trainer as feedback-giver. The literature offers some insights into this role but a great deal remains to be discovered about factors influencing feedback-giving behaviour. We studied factors within the personality of the trainer, hypothesizing that self-efficacy, task perception, neuroticism, extraversion, agreeableness and conscientiousness influenced the frequency, quality and the consequential impact of feedback. We found that 'task perception' and the personality trait 'neuroticism' affected the frequency and quality of feedback. Trainers who saw it as their task to organize and support the use of feedback gave more and better feedback, while trainers with a more neurotic personality observed trainees more frequently and gave better feedback compared to trainers with an emotionally more stable personality. A possible explanation may be that the more neurotic trainers felt insecure about leaving patients in the hands of trainees and

consequently observed more frequently. Because of the influence of trainers' task interpretation on feedback-giving behaviour, it seems worthwhile during training for trainers to pay attention to this aspect in addition to didactic skills which are usually the focus of faculty development in postgraduate medical education. Perhaps it is not so much shortcomings of didactic skills but rather trainers' task interpretation that determines their performance. It may therefore be useful to approach faculty development from a different angle.

In **Chapter 7** we return to the three main questions concerning the influence of the three elements, assessment instrument, trainee and trainer on observation of and feedback on single patient encounters. We conclude that in order to optimize learning through assessment, observation and feedback, we should target our efforts not at the format of the assessment instrument but at the way it is used by trainers and trainees. The crucial insights we gained in this respect can be summarized as follows: whilst the trainee influences the process, it is the trainer in particular who can activate and motivate trainees to achieve meaningful feedback that promotes learning. It should be noted, however, that the beneficial influence of the trainer on optimizing the feedback process depends crucially on how the trainer interprets his/her task. We relate our findings to models in the literature: Korthagen's onion model and the model of programmatic assessment, discussed in Chapter 1. After discussing the strengths and weaknesses of the research in this thesis, we present recommendations for further research and implications for practice. Considering the strong influence of the trainer on the feedback process we recommend further research on factors influencing feedback-giving behaviour. Further research should also investigate whether our results can be extrapolated to other medical specialties, undergraduate medial education or even totally different educational domains. Finally, we recommend research into training of trainers and trainees with regard to workplace-based assessment. With regard to the practical implications we specifically recommend to invest in the stakeholders of the feedback process.

In conclusion, our objective was to clarify observation and feedback in workplace-based learning and assessment by examining the role of the assessment instrument, the trainee and the trainer. We found that the process of observation of and feedback on trainee-patient encounters centres around the human actors: the trainee and even more importantly the trainer, while the assessment instrument is clearly of minor importance. Observation and feedback in workplace-based assessment were found to constitute a highly complex process with several actors and factors interacting in intricate patterns which determine whether observation takes place at all and, if it does, whether feedback is used to further the learning and development of trainees.

Samenvatting

Het verhelderen van observeren en evaluatieve feedback tijdens werkplekleren

In dit proefschrift onderzoeken we factoren die invloed hebben op toetsing door middel van feedback op geobserveerde patiëntencontacten bij werkplekleren. Het proefschrift is opgebouwd uit studies die tussen 2009 en 2012 zijn uitgevoerd binnen de huisartsopleidingen in Nederland. Na een algemene inleiding over de onderzoeksgebieden werkplekleren en toetsing worden in vijf hoofdstukken de resultaten van de verschillende onderzoeken besproken, gevolgd door een algemene discussie.

Hoofdstuk 1 markeert de directe aanleiding tot het proefschrift: een praktisch vraagstuk binnen de huisartsopleiding in Nederland. De aios-enquête liet al verschillende jaren op rij zien dat veel aios niet of nauwelijks geobserveerd worden tijdens patiëntencontacten in hun huisartsenstage. De literatuur laat ook zien dat studenten geneeskunde en aios in andere specialismen weinig geobserveerd worden tijdens het leren op de werkplek, wat er vervolgens toe leidt dat er weinig feedback gegeven wordt. In de inleiding bespreken we eerst de onderzoeksgebieden werkplekleren en toetsing. We beginnen breed en besteden aandacht aan theoretische perspectieven zoals de piramide van Miller, formatieve en summatieve toetsing en een model van programmatische toetsing. Vervolgens richten we onze aandacht of observatie en feedback. In dit hoofdstuk concluderen we dat het complexe terrein van werkplekleren en toetsing bestaat uit drie elementen: de *aios,* die zich moet ontwikkelen; de *opleider,* die dit leerproces begeleidt en evalueert; en een context met echte patiënten die niet volledig kan en ook niet moet worden gestandaardiseerd voor toetsingsdoeleinden. Wel dient er gezorgd te worden voor een eerlijke procedure die bruikbare feedback oplevert. Vermoedelijk kan dit verwezenlijkt worden met behulp van een *toetsingsinstrument.* Het doel van dit proefschrift is om te verhelderen welke rol observatie en feedback spelen bij leren en toetsing op de werkplek. Om dit proces in deze complexe context te ontrafelen bespreken we in de achtereenvolgende hoofdstukken toetsingsinstrumenten, de invloed van de aios en de invloed van de opleider.

Hoofdstuk 2 laat de resultaten zien van een kritische literatuurstudie over toetsingsinstrumenten voor klinische patiëntencontacten. Deze instrumenten kunnen aios inzicht geven in de sterke en zwakke punten van hun klinisch handelen. De resultaten laten zien dat de meeste instrumenten bedoeld zijn en gebruikt worden voor formatieve toetsing. Over het algemeen worden deze instrumenten in de literatuur goed ontvangen. De haalbaarheid wordt als goed beschouwd en met tien patiënten-

contacten wordt een acceptabele betrouwbaarheid bereikt. Van veel instrumenten is de validiteit niet onderzocht, maar de validiteit van de 'mini-CEX' en de 'clinical evaluation exercise' wordt ondersteund door een sterk en significant verband met andere gevalideerde toetsingsinstrumenten. De onderzochte literatuur levert echter geen overtuigend bewijs voor leereffecten. Geen van de studies onderzocht of het instrument leidde tot verbetering van leren, klinische vaardigheden of de kwaliteit van de patiëntenzorg. Ook is er een grote variatie in de beoordeelde competenties, het gebruik van beoordelingsschalen, het referentiekader en training van de gebruikers. De beoordelingskenmerken en de interpretatie daarvan blijven doorgaans impliciet en worden grotendeels overgelaten aan de opleiders. Training kan de sleutel zijn tot een doelmatige toepassing van toetsingsinstrumenten en het gebruik ervan verbeteren.

Hoofdstuk 3 behelst een onderzoek naar formele feedback naar aanleiding van een geobserveerd patiëntencontact. Het beoogde effect van deze feedback is dat de aios deze gebruiken om ervan te leren en hun prestaties te verbeteren. Dit is gebaseerd op de veronderstelling dat feedback leidt tot bewustwording van tekortkomingen en daardoor motiveert tot verandering en verbetering. Het onderzoek is gericht op externe factoren die van invloed zijn op feedback naar aanleiding van observaties tijdens werkplekleren. De resultaten van interviews met aios laten zien dat feedback betere resultaten oplevert als: 1) observatie en feedback door de aios en de opleider worden gepland; 2) de inhoud van de feedback en de manier waarop deze gegeven wordt doelmatig zijn; en 3) de aios de feedback aan leerdoelen koppelt om zijn of haar leren te sturen. De analyse van de interviews laat zien dat aios in eerste instantie emotioneel reageren op observatie, waarbij ze spontaan spreken over gevoelens van angst. Hoewel de emotionele reacties uniform is, hanteren aios verschillende strategieën om met deze emoties om te gaan. Sommigen vermijden observatie, maar anderen overwinnen hun angst en ondernemen stappen om ervoor te zorgen dat ze geobserveerd worden en feedback ontvangen. Vermijdingsgedrag van aios kan worden gecompenseerd door een actieve opleider. Gebaseerd op de data uit de interviews kunnen twee groepen opleiders worden onderscheiden: een groep die het initiatief neemt tot observatie met feedback en een groep die dit niet actief doet.

In **Hoofdstuk 4** beschrijven we een onderzoek naar verhalend commentaar naar aanleiding van geobserveerde patiëntencontacten bestaand uit feedback gegeven door opleiders, zelfreflecties van aios en actieplannen geformuleerd door opleider en aios samen. We onderzochten de hoeveelheid en specificiteit van deze commentaren. De commentaren werden ingevuld op een formulier ontworpen volgens algemene ideeën over feedback. Feedback is bij voorkeur geschreven en specifiek. Er dient duidelijk aangegeven te worden waar de ontvanger van de feedback aandacht aan

moet besteden. Feedback kan verbeterd worden door de ontvangers begeleiding te geven bij het omzetten van feedback in concrete stappen om prestaties te verbeteren. Feedback dient niet uitsluitend bepaald te worden door de opleider, maar er moet sprake te zijn van tweerichtingsverkeer waarbij de opleider feedback geeft en tegelijkertijd de aios aanmoedigt te reflecteren op zijn of haar prestatie. We onderzochten de effecten van dit formulier waarop duidelijk aangegeven werd waar opleiders en aios welk soort commentaar (feedback, zelfreflectie en een actieplan) konden invullen om zo het geven van commentaren te stimuleren. De resultaten laten zien dat feedback van opleiders vaker werd ingevuld dan zelfreflecties van aios. Er werden heel weinig actieplannen ingevuld. De meeste commentaren waren specifiek. Vanwege de verschillen tussen aios-opleiderkoppels in het aantal ingeleverde formuleren en omdat aios gedurende een jaar gekoppeld zijn aan één opleider, is de analyse gericht op verschillen tussen aios-opleiderkoppels. De resultaten laten grote verschillen zien. Sommige koppels geven consequent specifiek commentaar op alle ingevulde formulieren, terwijl andere koppels geen enkel specifiek commentaar invullen. Hoewel de opmaak van het formulier aangeeft welk type commentaar op welke plek moet worden ingevuld, behouden de gebruikers de mogelijkheid om het formulier op een andere manier te gebruiken. Op basis van de grote verschillen tussen de aios-opleiderkoppels concluderen we dat het feedback-proces niet zozeer wordt bepaald door het instrument, maar door de gebruikers. Hieruit kan worden geconcludeerd dat interventies om het leereffect van feedback te vergroten niet gericht moeten worden op het instrument, maar op de gebruikers ervan.

Hoofdstuk 5 bouwt voort op de resultaten van Hoofdstuk 4 met een onderzoek naar de relatie tussen reflectie, feedback en het formuleren van een actieplan. Ondanks de afwezigheid van empirisch bewijs, wordt er in de literatuur aangenomen dat (zelf) reflectie een positief effect heeft op het gebruik van feedback. In dit onderzoek onderzochten wij of feedback en reflectie een opbouwende reeks laten zien. Anders gezegd, slechts wanneer opleiders specifieke feedback geven, besteden zij aandacht aan de reflecties van hun aios. Ook veronderstelden wij dat reflectie een positieve uitwerking heeft op het gebruik van feedback, wat betekent dat aios die reflecteren op hun prestaties gekregen feedback vaker toepassen. Om deze aannames te toetsen hebben we de formulieren uit het onderzoek in Hoofdstuk 4 onderzocht. De resultaten bevestigen beide veronderstellingen. We vonden een opbouwende reeks van feedback en reflectie evenals een aanvullende werking van reflectie op het gebruik van feedback. Specifieke reflecties werden alleen gesignaleerd in combinatie met specifieke feedback en deze combinatie kwam veel vaker voor met specifieke actieplannen. Dit levert nieuw bewijs dat (zelf)reflectie het gebruik van ontvangen feedback kan stimuleren. Blijkbaar speelt een reflectieve houding ten opzichte van de

eigen prestaties een belangrijke rol bij het toepassen van feedback voor het bevorderen van het leren en het verbeteren van de eigen prestaties. De opbouwende reeks van feedback-reflectie-actieplan die wij vonden in dit hoofdstuk kan worden gebruikt bij ontwikkelingen op het gebied van feedback. Feedback krijgt al langere tijd veel aandacht in het (medisch) onderwijs en opleiders hebben al veel geleerd over het belang van feedback en methodes om feedback te geven. Het belang van reflectie en het verband tussen reflectie en feedback bij het vaststellen van leerdoelen is echter een latere ontwikkeling. Deze ideeën moeten hun weg nog vinden naar de standaard feedbackprocedures van veel opleiders.

Hoofdstuk 6 onderzoekt factoren die invloed hebben op de opleider in zijn of haar rol als feedbackgever. Hoewel de literatuur wel enig inzicht geeft in deze rol, is er niet veel bekend over factoren die van invloed zijn op gedrag met betrekking tot het geven van feedback. Het doel van dit hoofdstuk is om nieuwe inzichten te bieden ten aanzien van factoren met betrekking tot de persoon van de feedbackgever die het geven van feedback beïnvloeden. Wij veronderstellen dat kwaliteiten van de trainer, zoals zich bekwaam voelen en taakopvatting en persoonskenmerken als neuroticisme, extraversie, prettig zijn in de omgang en ordelijkheid van invloed zijn op drie elementen van het feedbackproces: frequentie, kwaliteit van de inhoud en effecten op het gebruik van de feedback. De resultaten laten zien dat taakopvatting en neuroticisme invloed hebben op de frequentie van de feedback en de kwaliteit van de inhoud. Op basis hiervan kunnen we concluderen dat een opleider die het als zijn of haar taak ziet om observatie en feedback te organiseren en het gebruik van feedback te bevorderen, meer en betere feedback geeft. Onze resultaten laten ook zien dat opleiders met een meer neurotische persoonlijkheid vaker observeren en betere feedback geven. Met andere woorden, emotioneel stabielere opleiders observeren minder vaak en geven minder goede feedback. Wellicht is er een verband tussen de frequentie en de kwaliteit van feedback en gevoelens van onzekerheid van meer neurotische personen, terwijl emotioneel stabielere personen dingen makkelijker uit handen geven. Gezien de relatie tussen taakopvatting en het geven van feedback, verdient het aanbeveling om in trainingen voor opleiders aandacht te besteden aan taakopvatting. Trainingen voor opleiders in de vervolgopleidingen richten zich veelal op (didactische) vaardigheden. Wellicht zijn het bij observatie en feedback niet deze vaardigheden maar is het eenvoudigweg de taakopvatting van de opleiders die maakt dat er niet wordt geobserveerd. Het verbeteren van de situatie zou dan vanuit een andere invalshoek bekeken kunnen worden.

In **Hoofdstuk 7** komen we terug op de drie hoofdvragen ten aanzien van de invloed van een toetsingsinstrument, de aios en de opleider op feedback naar aanleiding van geobserveerde patiëntencontacten. Het onderzoek leidt tot de conclusie dat niet het toetsingsinstrument leidend is bij het optimaliseren van leren door observatie en feedback, maar de manier waarop het instrument wordt gebruikt. De aios heeft zeker invloed, maar de opleider is bij uitstek degene die het proces van observatie en feedback kan activeren om te komen tot zinvolle feedback die een impuls kan geven aan de ontwikkeling van de aios. Hieraan dient toegevoegd te worden dat de opleider sneller initiatief zal nemen als hij of zij het organiseren van observatie en feedback als zijn of haar taak ziet. In dit hoofdstuk besteden we ook aandacht aan relaties tussen onze bevindingen en theoretische modellen uit de literatuur, zoals het ui-model van Korthagen en het model voor programmatische toetsing zoals besproken in hoofdstuk 1. In dit laatste hoofdstuk bespreken we ook de sterktes en zwaktes van het onderzoek in dit proefschrift. Verder doen wij aanbevelingen voor verder onderzoek en wijzen we op praktische implicaties van de bevindingen voor de onderwijspraktijk. Gezien de centrale rol van de opleider in het feedbackproces lijkt het zinvol om verder onderzoek te doen naar factoren met betrekking tot het geven van feedback. Ook is meer onderzoek nodig om na te gaan of onze resultaten van toepassing zijn op andere medische specialismen, geneeskundestudenten of zelfs andere onderwijsdomeinen. Tenslotte, verdient het aanbeveling om onderzoek te doen naar trainingen voor opleiders en aios. Met betrekking tot de praktische implicaties van de resultaten adviseren wij om specifiek te investeren in de belanghebbenden in het feedbackproces. Dit proefschrift had tot doel inzicht te bieden in observatie en feedback tijdens toetsing en leren op de werkplek. We hebben dit gedaan door middel van onderzoek naar de rol van het toetsingsinstrument, de aios en de opleider in het feedbackproces gebaseerd op observatie van patiëntencontacten. De uitkomsten wijzen op een centrale rol voor de menselijke actoren in het proces, de aios en in nog sterkere mate de opleider. Het toetsingsinstrument is van ondergeschikt belang. Het toetsingsproces op de werkplek door middel van feedback op basis van geobserveerde patiëntencontacten blijkt een belangrijk maar uitermate complex fenomeen, met verschillende actoren en factoren die elkaar beïnvloeden in ingewikkelde patronen die bepalen of observatie plaatsvindt, maar ook of gegeven feedback gebruikt wordt om het leren en de ontwikkeling van de aios te bevorderen.

Dankwoord

Met dit schrijven rond ik een periode in mijn leven af. Eigenlijk stond in deze periode 'leren' centraal. De afgelopen vier en een half jaar was promoveren eigenlijk de constante factor. De begrippen 'leren', 'toetsen', 'observatie', 'feedback' en 'reflectie' heb ik in theoretische zin uitgediept in dit proefschrift. Naast de drie vaste dagen die ik aan het onderzoek besteedde, heb ik in de praktijk deze begrippen verder betekenis gegeven door ondermeer te roeien, te coachen en een opleiding te doen tot leraar in het basisonderwijs. Dit alles samen heeft geleid tot een prachtige periode in mijn leven. Een heleboel dingen die ik in deze periode heb gedaan en mensen die ik heb mogen leren kennen, hebben direct of indirect een bijdrage geleverd aan het gereed komen van dit proefschrift. Hoewel ik weet dat wanneer je begint met mensen te bedanken, je er natuurlijk ook een aantal gaat vergeten, ga ik toch een poging doen.

Allereerst Cees van der Vleuten, mijn promotor. Zonder jou was dit onderzoek niet zo goed geworden als dat het nu is. Je corrigeerde mijn werk met een rode pen, maar altijd op een positieve en constructieve manier. Je gaf je feedback altijd binnen no-time. Ik hoefde het maar op te sturen en binnen de kortste keren had ik het nagekeken terug. Ik heb goede herinneringen aan onze gesprekken over het doen van onderzoek en de zin en onzin van de RCT versus kwalitatief onderzoek. Ook hadden we discussies over 'goede feedback'. Cees, ik wil jou danken voor alle goede, specifieke, formatieve, opbouwende, motiverende feedback die ik van je heb gekregen. Ik wens alle promovendi zo'n promotor zoals jij voor mij bent geweest.

Anneke Kramer, mijn copromotor heeft mij vanaf mijn eerste werkdag wegwijs gemaakt in het medische domein en meer specifiek de huisartsopleiding. Als enige arts in mijn promotiecommissie heb je goed je stem laten horen tussen al het geweld van die sociale wetenschappers. Hartelijk dank voor je goede inbreng en de ruimte die je me hebt gegeven om me te ontwikkelen.

Henk Mokkink, als copromotor had je een heel duidelijke inbreng in mijn kwantitatieve studies. Vooral het vijfde hoofdstuk en de daaruit voortkomende presentatie op het symposium in Bern is te danken aan jouw voorwerk, zelfs ondanks de gestolen computer en usb stick! Hartelijk bedankt voor je inzet en je wijze raad.

Vervolgens wil ik graag Mereke Gorsira bedanken voor het editten van mijn teksten. Het resultaat was altijd fantastisch en het wachten waard! Je hebt van mijn teksten goed lopend en lekker leesbaar Engels gemaakt.

Ben Bottema, ik wil jou bedanken voor het gestelde vertrouwen. In 2008 kwam ik als net afgestudeerde onderwijskundige, solliciteren op een functie waar jij eigenlijk een huisarts voor zocht. Hartelijk dank voor de kansen die je me hebt gegeven.

Zonder de bijdrage van participanten is het niet mogelijk om onderzoek te doen. Ik wil dan ook nadrukkelijk alle aios en opleiders in Groningen, Rotterdam, Utrecht, Maastricht en Nijmegen bedanken die hebben meegewerkt aan interviews, vragenlijsten en vooral voor het belangeloos inleveren van hun KKB-, KPB-, of KCT-formulieren.

Dankzij de contactpersonen op de verschillende locaties heeft het onderzoek daar kunnen plaatsvinden: Agnes Diemers, Margit Vermeulen, Frits Bareman, Bas Maiburg hartelijk dank hiervoor. Ook wil ik alle mensen die vanuit de secretariaten op de verschillende locaties een bijdrage hebben geleverd bedanken voor hun ondersteuning. Op deze plek wil ik ook Elke Butterbrod bedanken voor haar ondersteuning bij het interviewonderzoek.

Onderzoeken gaat lang niet zo goed wanneer je geen collega's hebt waar je even mee kunt sparren, je problemen mee kunt delen en nieuwe frisse ideeën bij kunt opdoen. Ik wil dan ook mijn ex-kamergenoten, ganggenoten, lunch-wandelingetjes-collega's, interview-samenwerkings-collega, alle ach-kun-je-me-even-in-Wijchen-afzetten-collega's en de collega's met wie we als 'jonge onderzoekers' etentjes organiseerden bedanken. Jullie weten nu zelf wel wie ik bedoel, maar toch: Fred, Thea, Patrick, Greetje, Nynke, Geurt, Bart, Anneke, Henk, Esther, Lia, Sanneke, Aggie en Marc, bedankt voor de fijne tijd!

Ook wil ik alle vrienden die ik de afgelopen jaren om mij heen heb gehad bedanken. Ze hebben me vooral veel ontspanning gebracht. Deze indirecte bijdrage heeft veel voor mij betekend. Naast alle drukke bezigheden is ontspanning echt belangrijk!

Speciaal wil ik noemen: Alle mensen die mij roeiend hebben ondersteund en waarmee ik zoveel plezier heb beleefd! In het bijzonder Jacoline en Joske (dank dat jullie mijn paranimfen willen zijn), AJ, Anke, Debbie, Marjolijn, maar zeker ook vele anderen! Mijn 'coachkindjes', in het bijzonder Vicky, Marloes, Geke en Simone. En natuurlijk mijn medecoach Suzanne.

Dan wil ik natuurlijk ook noemen: de dames van DRG (Dames RoeiGezelschap) Chimaera. In de beginjaren van mijn promotie was ik nog actief lid en hebben we vele gezellige borrels en weekenden doorgebracht. Special thanks voor mijn jaargenootje Floor, die net als ik na haar afstuderen is gaan promoveren. Zodat we nu als enige Chimaera-generatie kunnen zeggen dat we 'allemaal' gepromoveerd zijn! En ook speciale dank voor het Chimaera-wielerpeleton! Ons trainingsweekend in Limburg, de Amstel Goldrace, Limburgs mooiste (in de stromende regen), de ronde van Nijmegen en alle fietstochten in Duitsland en omstreken, dat er nog vele van dat soort tochten mogen volgen!

Tegenwoordig vind ik mijn ontspanning ook (weer) in het hockeyen. Hopelijk mogen er nog vele mooie jaren volgen met de gezellige dames van MHC Wijchen!

Tenslotte heel veel dank aan de vrienden die er al jaren voor mij zijn. Ontmoet op de middelbare school of in mijn studententijd, welke ik niet iedere week zie, maar met wie ik als we zo nu en dan afspreken hele gezellige, waardevolle dagjes 'als vanouds' heb: Ellen, Simone, Suzan, Maartje. Dat er nog vele gezellige uitjes mogen volgen.

Natuurlijk ook heel veel dank aan mijn familie.

Wim en Jessie, mensen uit de praktijk van het onderwijsveld. Wim, vanaf je lio-jaar had je een heel duidelijke mening over 'onderwijskundigen'. Lekker om daar over te discussiëren en grappig genoeg is Jessie nu ook een Master onderwijsweten-schap aan het doen. Erg leuk om met jullie te kletsen over 'leren' en 'goed onderwijs'. Jessie super bedankt voor het ontwerpen van de mooie omslag van dit proefschrift. Loes, die zelf niet houdt van leren en studeren, maar wel van 'kunnen' en 'meemaken'. Jij komt er wel! Berlijn klinkt wel ver weg, maar gelukkig is het ook heel dichtbij te halen met Skype, trein, auto en vliegtuig.

Natuurlijk pap en mam. Pap, van wie ik het 'alles willen weten' heb. Die ik (toen ik 6 was al) kon vragen naar de werking van hotelschakelingen: 'pap, als ik beneden op het knopje druk, waarom flapt het knopje boven dan niet de andere kant op als de lamp aan gaat'? Of later in mijn leven vragen stelde tijdens onze fiets- of schaats-tochten: 'Pap, waarom hebben ze eigenlijk de Noordoostpolder wel aan Nederland vastgemaakt en Zuidelijk Flevoland niet? Papa, fijn dat jij bijna altijd het antwoord wist, als ik iets wilde weten. En fijn dat je me gestimuleerd hebt om de antwoorden te gaan zoeken en te gaan vinden, op de dingen de ik niet wist en die jij me ook niet kon vertellen. En natuurlijk bedankt voor al die gezellige fietstochten en schaatspartijen! Mama, van jouw heb ik van kleins af aan geleerd om te voltooien waar ik aan begonnen ben (nee Els, als je hebt gekozen om te gaan kleien, gaan we niet na 5 minuten al weer opruimen en dan tekenen!). Een promotietraject is niet altijd leuk en soms krijg je de neiging om ermee op te houden. Maar doordat ik heb geleerd dat de voldoening heel groot kan zijn wanneer je afmaakt waar je aan bent begonnen, heb ik het kunnen voltooien. Ook heb ik van jou geleerd om dingen op te ruimen en netjes achter te laten voordat je aan iets nieuws begint. Zodoende ben ik de laatste dingen op het UMC nu dan ook netjes aan het afronden. Mam, dank voor al dit soort wijze lessen waar ik dagelijks de vruchten van pluk.

Last, but not least, Stef. Tijdens mijn promotietraject heb ik jou leren kennen en sindsdien hebben wij zoveel moois samen. Dat is eigenlijk niet onder woorden te brengen. We kunnen uren samen (na 'officiële werktijd') met de laptop op de bank naast elkaar zitten om de doelen te bereiken die we onszelf gesteld hebben. Maar

omdat we allebei af en toe zo gek zijn dat we maar doorgaan, weten we elkaar ook zo goed in balans te houden. Soms trek ik jou achter je computer vandaan en soms houd jij mij tegen als ik per sé weer eens 100 dingen in één week af wil hebben. Ook heb ik dankzij jou weer een heleboel nieuwe mooie, lieve mensen leren kennen! Schat, dankjewel voor alles, voor de fiets-, schaat- en wandeltochten waarbij we uren kunnen praten over van alles; al je hulp; alle keuze die we samen maakten; en alle mooie dagen die we samen al beleefd hebben. Dat we nog 80 jaar zo door mogen gaan!

Curriculum Vitae

Elisabeth (Els) Pelgrim is geboren op 7 november 1985 in Eindhoven als dochter van Rob Pelgrim en Lian Pelgrim-Sluijmer. Ze groeide op met haar broer (Wim) en zus (Loes) in het Twentse Borne. In 2004 behaalde ze haar middelbare school diploma (gymnasium) aan Lyceum de Grundel in Hengelo. In 2004 begon zij aan de studie Pedagogische Wetenschappen en Onderwijskunde aan de Radboud Universiteit Nijmegen. In 2008 behaalde ze haar Master Onderwijskunde (Bene Meritum). In het Master jaar liep ze stage en schreef ze haar scriptie op de onderwijsafdeling van het UMC St Radboud in Nijmegen. Direct na haar afstuderen begon ze in oktober 2008 met het promotieonderzoek naar het observeren en het geven van evaluatieve feedback tijdens werkplekleren van huisartsen in opleiding bij de afdeling Eerstelijns-geneeskunde van het UMC St Radboud, wat resulteerde in de publicatie van dit proefschrift.

Naast haar studie onderwijskunde en in de eerste jaren van de promotie roeide zij op nationaal niveau bij de Nijmeegse Studenten Roeivereniging Phocas, waarmee zij meerdere nationale en internationale medailles behaalde. In 2010 en 2011 werkte zij naast haar promotietraject als junior onderzoeker bij de afdeling Eerstelijnsge-neeskunde aan de ontwikkeling van de hao-aios toets en een in samenwerking met het Nederlands Huisarts Genootschap (NHG) ontwikkeld voorschrift voor de basisuit-rusting van huisartspraktijken en huisarts opleidingspraktijken. Sinds 2011 is zij, naast haar promotietraject, in opleiding tot lerares in het basisonderwijs aan de Fontys Pabo Eindhoven.

Els woont samen met haar vriend Stef Roskam in Wijchen.

Curriculum Vitae

Curriculum Vitae

Elisabeth (Els) Pelgrim was born on November 7th 1985 in Eindhoven (The Netherlands) as daughter of Rob Pelgrim and Lian Pelgrim-Sluijmer. She grew up with her brother (Wim) and sister (Loes) in Borne. She completed her high school (gymnasium) in 2004 at Lyceum de Grundel in Hengelo. And in the same year she started her study Pedagogical en Educational Science at the Radboud University in Nijmegen. In 2008 she graduated for her master's degree in Educational Science with the laudation Bene Meritum. In the last year of her study she did her internship and wrote her master thesis at the educational department of the Radboud University Nijmegen Medical Centre. Directly after her graduation, in October 2008, she started her PhD into observation and assessment feedback in workplace-based learning of general practice trainees at the department of primary care and community care at the Radboud University Nijmegen Medical Centre. This resulted in the publication of this thesis.

In addition to her study Educational Science and in the first years of her PhD, she was a rowing athlete and member of the student rowing club of Nijmegen (Phocas), for which she won several national and international medals. In 2010 and 2011 she worked as a junior researcher at the department of primary care and community care in addition to her PhD project. The project she worked on contained the development of a method of assessment in which GP-trainees review their trainer. And in cooperation with the Dutch College of General Practitioners (NHG) she developed guidelines for the basic equipment of regular general practices and training practices. Since 2011, she is training to be a primary school teacher at the University of Applied Sciences Fontys in Eindhoven.

Els happily lives with her friend Stef Roskam in Wijchen.

172