Patient Feedback in General Practice Training
The studies presented in this thesis were conducted within the EMGO+ Institute (www.emgo.nl). The EMGO+ Institute participates in the Netherlands School of Primary Care Research (CaRe) which was re-acknowledged in 2005 by the Royal Netherlands Academy of Arts and Sciences.

The studies presented in this thesis were financially supported by the Dutch Foundation for the Vocational Training of General Practice.

Financial support for the printing of this thesis has kindly been provided by the Dutch Foundation for the Vocational Training of General Practice, by the EMGO+ Institute, and by the VU University.


Layout: Krista Miedema
Printed by: Gildeprint Drukkerijen – www.gildeprint.nl

© M.E. Reinders, Haarlem, the Netherlands, 2010. All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system, without prior permission of the holder of the copyright.
Patient Feedback in General Practice Training

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Geneeskunde
op dinsdag 18 mei 2010 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

doors

Marcel Egge Reinders

geboren te Amsterdam
promotor: prof.dr. H.E. van der Horst

copromotoren: dr. A.H. Blankenstein
dr. H.W.J. van Marwijk
**Contents**

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>Development and feasibility</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>Validity of the questionnaire</td>
<td>39</td>
</tr>
<tr>
<td>4</td>
<td>Patient view on patient-centeredness</td>
<td>57</td>
</tr>
<tr>
<td>5</td>
<td>Effectiveness (controlled trial)</td>
<td>73</td>
</tr>
<tr>
<td>6</td>
<td>Context of existing literature (systematic review)</td>
<td>95</td>
</tr>
<tr>
<td>7</td>
<td>GPT participation</td>
<td>127</td>
</tr>
<tr>
<td>8</td>
<td>Generalisability</td>
<td>139</td>
</tr>
<tr>
<td>9</td>
<td>General discussion</td>
<td>157</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td>183</td>
</tr>
<tr>
<td></td>
<td>Samenvatting</td>
<td>189</td>
</tr>
<tr>
<td></td>
<td>Nawoord</td>
<td>197</td>
</tr>
<tr>
<td>10</td>
<td>About the author</td>
<td>203</td>
</tr>
</tbody>
</table>
 CHAPTER 1

Introduction
The need for adequately communicating doctors is widely accepted and is related to a plethora of patient related health outcomes\(^1\). General practice vocational training institutes and the general practitioner trainers can help general practice (GP) trainees to learn the necessary communication skills, by developing educational materials for effective teaching\(^2\). In the medical field, and particularly in general practice, there has been a change from focusing more on the patient rather than on the disease\(^3\). From this perspective, patient involvement and a patient-centered approach may not only be helpful in the process of learning consultation skills, they are in fact essential. We now need to know how educational interventions that enhance the involvement of real patients can best be evaluated and applied.

In this thesis we describe the steps that are required to build up evidence for patient feedback as an educational intervention, which is complex and multifaceted by nature.

**Context of the study**

The perception of adequate and effective communication between patients and doctors is subjected to change over time: from of a well-defined, well-delivered message communication style on the part of the doctor to a more interactive, give and take approach\(^4\). The goal should be set for a sender and a receiver reciprocal agreement. Effective communication is correlated with patient-related outcomes: improved health status, reduction of non-evidence-based prescriptions, reduced litigations, and patient satisfaction. It is self-evident that patients need adequate communication with doctors, but it is not only the patients who benefit from the good consultation skills of a general practitioner; doctors themselves can reciprocally benefit, for example by reduced diagnostic testing\(^5\).

It cannot be said that education on communication in general practice education has always been an underprivileged area. On the contrary, communication skills have traditionally (and rightly) been seen as one of the cornerstones of primary care. To set goals for graduating general practitioners, a competency profile was defined, in which the key competency ‘communicator’ covers the quality items of doctor-patient communication\(^6\).
Less well-defined and less known, however, is how GP trainees (or doctors in general) acquire good medical communication skills. Is this a natural process stimulated by ‘learning-by-doing’, or can it be influenced by rigorous methods of education? Assessment of significant improvements in communication performance, as a result of educational interventions or even systematic teaching, is notoriously difficult\(^7\), possibly due to a lack of sensitive instruments and a lack of outcome definitions.

In a recent framework for the curricula of the vocational training for general practitioners it is stated that ‘the scientific base for primary care medicine has advanced quickly, with an exponential increase of acquired knowledge and many guidelines as a result’. It is therefore remarkable that there is a great diversity in guidelines for doctor-patient communication, but very little evidence\(^8\). In fact, the implementation of more uniform teaching methods among the vocational training institutes is more justifiable when their effects have been proven.

The question that arises is to what extent there is evidence for medical teaching methods. Many studies have focused on the evaluation of programmes or satisfaction surveys, and these studies are meant to provide information for a wider audience\(^9\). However, the impact on evidence-based teaching is limited, especially if there is a lack of rigorous standards for the validity and reliability of methods and instruments.

What is needed is research to evaluate the effectiveness of educational interventions, which will make it possible to achieve higher levels of evidence-based teaching. Unfortunately, health professionals are often reluctant to participate in research on the effectiveness of educational interventions\(^10\). Opinions differ, not only with regard to the areas of need, but also about what ‘works’, and with which groups. Furthermore, the methodology of educational interventions is severely hampered by their complex nature, the difficulty of including a sufficiently large number of participants, and the difficulty of assessing actual change as an outcome measure\(^7,10\).
There are several models that can be used to conceptualise levels of learning evidence. One basic model, developed by Kirkpatrick in 1967\textsuperscript{11}, has four levels of educational impact which measure:

- the reaction of students; what they thought and felt about the training
- learning; the resulting increase in knowledge or capability
- behaviour; the extent of improvement in behaviour
- implementation/application; the achieved change in the performance of the trainees.

Each level aims at a different level of educational impact, and therefore a different level of evidence-based teaching. Methodologically, the assessment of the level of evidence also greatly depends on the rigorousness of the applied methods and measurement instruments.

So, although much effort is put into teaching consultation skills, it is not clear how effective this is. The question that remains is: how can we further stimulate the learning of adequate communication skills effectively, within a student-centered context? Therefore, the staff of the vocational institute for general practice and the research staff worked together to develop the patient feedback programme. Patients are an important source from which to learn about consultation skills; it is simply a matter of how to optimize this learning potential, and how to evaluate it. First-year GP trainees, who were in the process of learning patient-doctor communication skills, participated in the patient feedback programme, and applied it in their daily practice. They were both the ‘objects’ and ‘subjects’ of study.

**Rationale of the study**

The choice for a patient feedback intervention is based on our wish to put greater emphasis on two major components of communication training that are believed to be effective for developing consultation skills: a greater involvement of patients in the education of GP trainees, and the development of new approaches to structured feedback assessment.

Real patient involvement might provide new stimuli in the patient-centered consultation style, which is already included to some extent in our vocational training for general practice. Patient-centeredness means that
physicians incorporate the opinions, experiences, wishes and concerns of patients in medical decision-making, within the limits of their own professional responsibilities\textsuperscript{3}. Patient-centered consultation skills have been found not only to enhance patient satisfaction, but also to improve various important outcomes, such as compliance to treatment, better health status and less symptom burden\textsuperscript{12}.

Feedback on their performance is highly valued by students\textsuperscript{13} (GP trainees) as well as by GP trainers and teachers. GP trainees find feedback useful, instructive, and reassuring. Furthermore, they value their teachers and GP trainers more if they make use of feedback as an educational tool. The effects of assessment and feedback on the performance of a doctor are probably greater when supported by an authoritative source, over an extended period of time\textsuperscript{12}. In theory, patients are in a good, or maybe the best, position to give feedback on the consultation skills of doctors, just because they are there, in sufficiently large numbers, and because they are the actual receivers of the care. However, intervention studies with patient feedback should ensure that patients are enabled to provide feedback according to well-defined criteria for safe and effective feedback\textsuperscript{2}. Good feedback has to meet several requirements. Most important of all is that the patients should be assured that giving feedback has no repercussions whatsoever on the care they receive or on their relationship with the doctor. Probably due to the unequal balance of power between doctors and patients, socially desirable feedback is a commonly encountered phenomenon, which limits the learning potential\textsuperscript{14}.

The GP trainee should also be well-prepared and comfortable with the feedback of patient assessments, in order to successfully adhere to such a programme. Important questions are: will GP trainees identify their needs for patient-centered consultation skills, and if so, how? Will they put these on their learning agenda, will they be susceptible for patient assessments and feedback, and most importantly, will they learn from the programme?
Objectives and outline of this thesis

Our aim of the studies described in this thesis is to evaluate the implementation of a patient feedback programme in the vocational training for general practice. Educational interventions of this type are complex projects, because the participants who are involved (including GP trainees, teachers and patients) all have different needs, priorities and personalities, many factors can influence the effectiveness of the interventions, and reliable and valid measurements instruments are scarce. That is why we chose a combination of various quantitative and qualitative research methods: a feasibility study (including qualitative research), a theoretical development study, validity studies, a controlled trial (in which the effectiveness of the patient feedback on the improvement of GP trainees was investigated), and a systematic review.

Development and feasibility (Chapter 2)

We piloted the patient feedback programme among first-year GP trainees, in a feasibility study in 2005. The trainees, their patients, their GP trainers, and their teachers at the institute were asked to give their opinions about the design and content of the programme. The qualitative part of this research tested the acceptance of patient feedback by the GP trainees, and we used the results to improve the feasibility of the programme.

Validity of the questionnaire (Chapter 3)

In this chapter we focussed on the validity of the patient feedback questionnaire on consultation skills (PFC), which is based on an existing questionnaire, the Patient Perception of Patient-Centeredness (PPPC). The PFC is an extension of the PPPC, including questions to cover all items of the key competency ‘communicator’. Patients and GP trainees were involved in the assessment of face validity, and experts were asked to assess the content validity. Subsequently we calculated clinimetric characteristics.

Patient view on patient-centeredness (Chapter 4)

We studied the patients’ perspective on components of patient centeredness in real consultations. Therefore patients were asked to complete a questionnaire,
the Patient Perception of Patient-centeredness (PPPC). Itemizing patient-centeredness into components (exploring both the disease and the illness experience, finding common ground, and the personal context) will provide more detailed information about the specific parts of a consultation that might need more attention.

**Effectiveness (controlled trial, Chapter 5)**

In a controlled trial, in which first-year GP trainees were randomly allocated to an intervention or a control group, we studied the actual effect: (improved) consultation skills as a result of patient feedback. The intervention group attended a patient feedback training programme, in addition to the regular communication training, whereas the control group received only the regular communication training. Standardised simulated patients visited the practices and video-taped the consultations at baseline and after three months. The consultations were assessed by eight trained staff-members.

**Context of existing literature (systematic review, Chapter 6)**

We conducted a systematic review to place our results in a wider context. In the systematic review we searched for evidence of the educational potential of real patient feedback on the general consultation skills of practicing physicians. Empirical studies with randomized (controlled) and various non-randomized designs (including qualitative research) were included, as long as the educational impact of patient feedback was described (varying from the evaluation of patient feedback to the assessment of change in outcomes).

**GPT participation (Chapter 7)**

The wide variation in adherence to the patient feedback programme by the GP trainees, as found in the trial study, was further investigated. We interviewed the GP trainees with a low level of adherence personally. We examined whether baseline consultation skills, such as performance on consultation skills, were correlated with the participation rate.
Generalisability (Chapter 8)
The teachers of the general practice vocational training of the VU medical center attended a short training course on the assessment of consultation skills with the MAAS-Global Instrument. They then assessed the video-taped recordings of consultations with real patients and simulated patients. All consultations were scored independently by two teachers, according to a randomized allocation procedure, providing results from which reliability and generalisability calculations could be made. With an evaluative questionnaire we assessed the perceived competence of the teachers after having attended the course.

General discussion (Chapter 9)
In this chapter we discussed the synthesis of the findings of the studies described in this thesis and we provided a critical reflection on what our findings add to the existing literature. We also discussed the strengths and weaknesses of the studies, and the implications for general practice vocational training and further research.

The thesis concludes with a summary of the design and results of the studies in both English and Dutch.
References


CHAPTER 2

Development and feasibility of a patient feedback programme to improve consultation skills in general practice training

Marcel E. Reinders
Annette H. Blankenstein
Harm W.J. van Marwijk
Harry Schleypen
Piet L. Schoonheim
Wim A.B. Stalman

Abstract

Objective
To develop an attractive and effective patient feedback training programme for general practice trainees (GPTs).

Methods
First, an exploratory study was conducted in which patients and GPTs were interviewed after they had worked with patient feedback. This contributed to the development of the patient feedback training programme. Subsequently, in a feasibility study, first-year GPTs asked patients to give feedback on their consultation skills by completing a questionnaire. The outcomes of group discussions with the GPTs and the results of the evaluation forms filled in by the GPTs were analysed.

Results
Forty-eight GPTs collected 878 questionnaires. GPTs and patients alike expected patient feedback to be a major tool for acquiring consultation skills. The GPTs encountered several obstacles in the organisation of this programme in their practice. They reported that the learning effects were more limited than they had expected because patients gave positively biased answers and because not all consultations provided an appropriate source of patient feedback.

Conclusion
The new patient feedback programme on consultation skills is feasible for patients and GPTs.

Practice Implications
To optimise the educational potential and benefits of patient feedback, GPTs should ask for feedback from patients after challenging consultations, and should stimulate patients to be critical in their answers.
Introduction

Adequate communication is essential for the provision of good patient care. Better understanding between doctors and patients results in greater patient satisfaction, higher compliance rates and improved health outcomes\textsuperscript{1,2}. Doctors could improve their consultation skills by adopting a more patient-centered approach\textsuperscript{3,4}, and by assessing the effects of their behaviour by asking patients for feedback.

In recent years the general practitioner vocational training institutes in the Netherlands have redefined a competency profile (based on the CanMeds Framework)\textsuperscript{5} for the post-graduate general practice trainees (GPTs) and restructured the framework of their programmes. As presented in the original CanMeds Framework, the Dutch competency profile also contains seven key competencies. The key competency “communicator” has served as a guiding principle for the communication training programmes. It has been emphasized that acquiring knowledge and skills is an active personal process, influenced by individual learning styles.

The current doctor-patient communication training programmes for GPTs are based on a patient-centered approach, the benefits of which have been demonstrated by Henbest and Stewart\textsuperscript{6} and confirmed by others\textsuperscript{7-11}. Another important aspect of the communication programmes is giving and receiving feedback, for which various educational tools are available. GPTs currently receive feedback from their GP trainers, their colleagues and the staff of their training institutes. However, systematic feedback from the target group, i.e. the patients, has not yet been implemented, although the GPTs do feel the need for it\textsuperscript{12}. An indirect comment from a patient, reported by the practice assistant or the GP trainer, is often not sufficient, and this is why it is important that they receive feedback directly from the patients.

In general, studies that have investigated the effects of educational interventions to improve the consultation skills of GPTs have reported conflicting results\textsuperscript{13-17}. Several studies have focused on the effect of patient feedback on improvement of health care in general and on patient satisfaction\textsuperscript{18}. The limited
number of studies that investigated the effectiveness of implementing patient feedback on the improvement of the consultation skills of general practitioners (GPs) or GPTs, has not provided univocal results\textsuperscript{19-22}. Improvement in the communication behaviour of GPs as a result of patient feedback could not be demonstrated for various reasons\textsuperscript{20,21}. However, one study has reported a positive effect of patient feedback on the interpersonal skills of GPTs\textsuperscript{19}.

Because of this lack of relevant studies and the contradictory results, we decided to develop a patient feedback training programme and to study its effect on the consultation skills of GPTs. We designed the programme in such way that it could be incorporated in our doctor-patient communication training programme, based on the following principles. First, the programme was designed as an educational intervention that was intended to improve the communication skills of GPTs, and not merely to improve patient satisfaction. The patient feedback would not be used to assess the communication skills of GPTs, but to provide the GPTs more or better insight into their (patient-centered) consultation skills. Furthermore, in view of the recently redefined competency profile, we considered it important that the method used (a questionnaire) would reflect the main requirements for the key competency of ‘communicator’. Moreover, in order to enhance the educational impact, we opted for the use of a questionnaire with a patient version (feedback) and a doctor version (self-assessment). The combination of these principles in a patient feedback training programme is a new concept, especially in the Netherlands.

The aim of this study was to give a qualitative description of the development and introduction of a patient feedback training programme for GPTs. In this paper we report on an exploratory study and a feasibility study. The following research questions were addressed:

- Is patient feedback by means of a questionnaire attainable for patients and GPTs in the Netherlands?
- How do GPTs react to the implementation of a new patient feedback programme?
- What changes need to be made in the programme?

The first research question was addressed in the exploratory study, and the second and third questions were addressed in the feasibility study.
Methods

The research was carried out in the GP vocational training institute of the VU medical center in Amsterdam. In order to answer the first research question we first carried out an exploratory study, in which third-year GPTs and patients were interviewed after they had been working with patient feedback. This contributed to the further development of the patient feedback programme (see box). To answer the second and third question we subsequently carried out a feasibility study, in which first-year GPTs worked with patient feedback after they had completed the patient feedback training programme (Figure 1). We assessed the appreciation and relevance of the patient feedback training programme, and also the response of the GPTs to the programme.

Exploratory study and development of the patient feedback programme

In July and August 2005 we carried out an exploratory study, in which patients, GPTs and GP trainers assessed the practicalities and appreciation of patient feedback. They made use of the modified Patient Perception of Patient Centeredness (PPPC) questionnaire, developed by Stewart et al\textsuperscript{23}, which correlates well with items of the key competency communicator. Patients can indicate their ratings on a 4-point scale. The PPPC is a one-dimensional instrument (tested on validity and reliability), that measures patient-centered communication aspects within the consultation. It has been used for research, as well as for education, by providing individual feedback to participating physicians on their patients’ perception. The definition of ‘Patient-Centeredness’ has been defined in several ways. The components of relevance have been described as: 1. exploring the disease and the illness experience, 2. understanding the whole person, 3. finding common ground, 4. incorporating prevention and health promotion, 5. enhancing the doctor-patient relationship, and 6. being realistic in timing and use of resources\textsuperscript{23}.

Our modifications consisted of translation into Dutch, slight simplifications of questions that patients indicated as difficult or unclear, and the addition of questions to cover all the items of the key competency.
Communicator\textsuperscript{5}, according to the advice of a panel of experts. We have assessed the face and content validity within the process of refinement of the questionnaire. The questionnaire has two parallel versions, one for the patient and one for the doctor. Assessment of the self-rating of the GPTs was beyond the scope of this article. Seventeen third-year GPTs participated voluntarily, asking patients at random to provide feedback on their consultation skills. One of the researchers was present in the waiting room of the practice, to hand out written information and to observe the process closely. Semi-structured evaluative interviews were held in the practices with the patients, the GPTs and their GP trainers.

**Feasibility study**

From September 2005 until February 2006 a patient feedback training programme was organised for all 58 first-year GPTs (divided into five groups), on days that they were present at our institute for educational purposes. Of the 58 GPTs who started their vocational training programme in March 2005, 10 did not complete the Patient Feedback Training Programme for the following reasons: pregnancy leave (4), a different educational trajectory (2) and general negative assessments (4).

The GPTs used the modified Patient Perception of Patient Centeredness (PPPC) questionnaire. A number of questions were added to this list, selected and further refined on the basis of the experiences from the exploratory study, to cover all items of the key competency ‘communicator’ (Table 1). The GPTs were advised to hand out 30 questionnaires, which we considered to be sufficient for a relevant learning experience, but not a burden.

All patient feedback questionnaires (doctor version) had an attachment containing questions about the application (Table 2). After they had been used for educational purposes, the patient feedback questionnaires and the attachments were collected by the teaching staff and analysed by the researchers. The results of the scoring on the questions of the patient feedback questionnaire will be analysed and presented separately.

Group interviews were held during the instruction meetings and evaluation meetings. One of the researchers was present in order to observe
and monitor the process, and to make notes of the most frequently mentioned comments. At the evaluation meeting all the GPTs filled in an evaluation form. Table 3 shows the questions (and the answer categories) of the evaluation form. All questions were put in an open format; the answer categories were identified according to the frequency of repeatedly stated answers.

### The Patient Feedback Training Programme

The training course consisted of (1) two instruction meetings, (2) patient feedback related assignments, and (3) an evaluation meeting.

1. At the *instruction meetings*, the concept of patient feedback was explained and the patient feedback questionnaire was demonstrated. The GPTs were informed that parallel to the educational project, a feasibility study would be carried out. The organisational aspects of patient feedback in their practice were discussed. In a setting with a simulated patient (actor), the GPTs could practice in asking patients to give feedback. In order to find out what would work best in their own practice, the GPTs were stimulated to discuss how, at what moment, and which patients they would ask to give feedback. The GPTs were referred to a website, on which all the necessary information could be found. A comparable course was given to the GP trainers. They were encouraged to coach their GPTs and to have regular topic discussions.

2. *Assignments*. The main assignment was that the GPTs were asked to hand out the questionnaire to 30 patients. There were no exclusion criteria, but we advised the GPTs not to involve patients with insufficient command of the Dutch language. We recommended them to hand out 15 forms at random (for example the first few consecutive patients in a consultation hour), and another 15 during the more challenging consultations (for example complex medical problems, unsatisfactory consultations). The patients filled in the questionnaires at the practice, directly after consulting the GPT. They then handed over the questionnaire in an envelope to the practice assistant. Although it did not contain the name of the patient, it could be traced to a certain patient by date of birth, and was therefore not anonymous. Despite the obvious disadvantage, this was necessary to link the two parallel versions. In order to enhance the learning effect, the GPTs were advised to study the patient questionnaires, make a comparison with their own version, and to formulate learning aspects, all on the same day. Related assignments were that the GPTs were instructed to make a video-recording of a consultation. The patient and the GPT filled in their own questionnaire, and the outcomes of both versions and the video were discussed with the GP trainer (360 degree angle perspective). The GPTs were also asked to analyse the cumulative of the given feedback, to compare it with their own version and to try to formulate learning aspects that may help to improve consultation skills. These results could be used as a subject of discussion with their GP trainers. During the months that the GPTs actually worked with patient feedback, the teachers and the researchers monitored their progress, and when necessary, encouraged the GPTs to hand out the questionnaires.

3. At the end of the project, an *evaluation meeting* was held. Relevant and frequently mentioned comments were used for further development of the programme.
Figure 1. Flow chart

Table 1. Patient Feedback Checklist

1. To what extent was/were your main problem(s) discussed today?
2. How satisfied were you with the discussion of your problem?
3. To what extent did the doctor listen to what you had to say?
4. To what extent did the doctor explain this problem to you?
5. To what extent did you and the doctor discuss your respective roles?
6. To what extent did the doctor explain treatment?
7. To what extent did the doctor explore how manageable this (treatment) would be for you?
8. How well do you think your doctor understood you today?
9. To what extent did the doctor discuss personal or family issues that might effect your health?
10. Did the doctor encourage you to ask questions?
11. Have you acquired more or new insight as a result of your visit to the doctor?
12. Do you have more insight into what you can do about your problem?
13. How well did the doctor help you today?
14. Was the doctor clear about the proposed plan(s)?

(Q 1-9 from: Stewart et al. Patient Perception of Patient Centeredness, Q 10-14 added as a result of patients’, doctors’ and experts opinions; answer scale: Completely-Mostly-A little-Not at all)
Table 2. Aspects of application of the checklists (48 GPTs, n = 878)

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who asked the patient for feedback?</td>
<td>GPT</td>
<td>85.7%</td>
</tr>
<tr>
<td></td>
<td>Practice Assistant</td>
<td>14.2%</td>
</tr>
<tr>
<td></td>
<td>Colleague</td>
<td>0.1%</td>
</tr>
<tr>
<td>When was the patient asked?</td>
<td>In the waiting room</td>
<td>16.8%</td>
</tr>
<tr>
<td></td>
<td>At the beginning of the consultation</td>
<td>26.5%</td>
</tr>
<tr>
<td></td>
<td>At the end of the consultation</td>
<td>56.7%</td>
</tr>
<tr>
<td>Which patients were asked?</td>
<td>At random</td>
<td>84.5%</td>
</tr>
<tr>
<td></td>
<td>For selected reasons</td>
<td>15.5%</td>
</tr>
<tr>
<td>To whom was the checklist returned?</td>
<td>GPT</td>
<td>11.4%</td>
</tr>
<tr>
<td></td>
<td>Practice Assistant</td>
<td>85.2%</td>
</tr>
<tr>
<td></td>
<td>Mailbox</td>
<td>3.5%</td>
</tr>
<tr>
<td>How did the consultation go?</td>
<td>Well</td>
<td>54.3%</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>39.7%</td>
</tr>
<tr>
<td></td>
<td>Difficult</td>
<td>6.1%</td>
</tr>
<tr>
<td>Was there a language barrier?</td>
<td>Yes</td>
<td>1.2%</td>
</tr>
<tr>
<td></td>
<td>Some</td>
<td>3.0%</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>95.8%</td>
</tr>
</tbody>
</table>
Results

Exploratory study

*Applicability of patient feedback and the questionnaire according to the patients*

Only two out of 71 patients were unable to participate (one because of pain, one because of lack of time). Two of the remaining 69 patients needed help in completing the questionnaire, because they had difficulty in understanding the questions. The mean age of the patients was 43.9 years (SD: 22.3, range: 2 – 82), 70.3% were female, and 42.4% had visited the doctor in question for the first time. According to the patients, their problem was simple (51.5%), or moderately complex (27.3%), or complicated (21.2%). The majority of the patients managed to fill in the questionnaire in 3-4 minutes and only three patients took longer than 5 minutes. None of them considered that it took too long.

Semi-structured interviews with the patients (n=51) revealed that they were generally positive to very positive (n=49) about the idea of giving feedback to the GPTs; two were only moderately positive, but none were negative. They expected patient feedback to benefit both doctor and patient, and were generally glad to make a contribution to the GPT training programme. The concept that their feedback would be used to improve the consultation skills of GPTs was clear for most of the patients, although some still thought that it was to assess patient satisfaction. Only four patients admitted that they found it difficult to give negative feedback to their doctor (especially if it was not anonymous). If they had been asked to fill in the questionnaires at home, none of the patients thought that they would have answered differently. They thought that the general content of the questions reflected the communicational aspects well. However, a number of patients felt that patient feedback was not always relevant when dealing with relatively simple medical problems.

*Applicability of patient feedback and the questionnaire according to the GPTs*

Semi-structured interviews were held with the GPTs (n=17) and their trainers. The GPTs considered patient feedback as a self-directed learning method, with
great potential for the improvement of communication skills. They completed the questionnaires in 1-3 minutes. Four GPTs said that they had acted somewhat differently during their consultation, in particular by putting more effort into the consultation and in asking more questions. Like the patients, the GPTs considered that the questionnaire reflected the communicational aspects of consultation, and they indicated which questions they thought were less relevant when patients presented relatively simple medical problems. One GPT mentioned that simply filling in the doctor version, or even the presence of the checklists on the desk, would enhance alertness, and would therefore benefit the consultation. Fourteen GPTs expected that patient feedback would provide enough educational material for discussion with their GP trainers. The GP trainers generally supported the concept, but mentioned that not all practices had a room in which the patients could fill in their questionnaires in privacy, which might be an organisational obstacle.

**Feasibility study**

**Expectations**

From the data out of the instruction meetings it appeared that the GPTs were generally enthusiastic about the idea of patient feedback, and they certainly felt a need for it. They were eager to know what the patients actually thought about their consultation skills, but some feared unreasonable criticism. On the other hand, the GPTs thought that patients would give socially acceptable answers if they had to fill in and return the list in the practice and could not do so at home. Some of the GPTs thought they would feel vulnerable in asking patient to give feedback, but none of them thought that patient feedback could have any negative repercussions on the assessments made by the GP trainers. However, the GPTs mentioned that they already had a full agenda, and that a project like this would certainly be an additional burden.

**Response**

Figure 2 shows the number of questionnaires that each GPT received from patients. A total of 878 questionnaires were received (med. 18, SD 8.6, range 0-32) by 48 GPTs. Only 22.9% of the GPTs actually completed the task (received
more than 25 questionnaires), 66.7% received 15 or more, and 8.3% received less than five questionnaires. Although the programme was obligatory, it appeared that two GPTs did not use the questionnaires at all. The GPTs who completed more than 15 appreciated the project more (p = 0.04, two-sided t-test).

Self-reported learning topics were formulated by 34 of the 48 GPTs (70.8%) and reported in an open format. Frequently reported topics were: to check if the given information to the patient is well understood, to ask what the patient really came for, to give more information to the patient, and to gain some confidence because the patient is more satisfied with the consultation than the doctor is.

Table 2 presents aspects of application of the checklists. A minority of 15.5% of the questionnaires were handed out to selected patients.

...Although I intended that morning to hand out the questionnaires after difficult consultations or to demanding patients...I ended up asking for feedback from patients who were friendly and cooperative...

As shown in Table 3, 37.0% of the GPTs completed the assignments and had two related topic discussions with their GP trainer, 32.6% had one discussion, and 30.4% had no discussions. Of the GPTs who did complete one or two assignments, the topic discussions with their GP trainers were considered useful by 47.6%.

**Time-burden**

In table 3, it is shown that the majority of 53.6% of the GPTs needed 5 minutes for instructing a patient (mean 5.6 minutes, range 1.5 –10); GPTs did not consider this to be too long. However, 78.3% of the GPTs thought that having to do this 30 times was a real burden, and only 15.2% had no problems in doing so. Interference of the consulting hours by acquiring patient feedback was mentioned by 26.5% of the GPTs.

**Appreciation**

A total of 46 evaluation forms were returned (93.8%). The majority of the GPTs (82.0%) were enthusiastic at the start, but at the end of the programme only
45.6% thought it would be beneficial for their education. The reasons GPTs mentioned for the limited usefulness are summarized in Table 3. The majority of the GPTs (67.4%) encountered various obstacles in organising the patient feedback programme in their practice.

The obstacles and reasons for adjusted learning options were also addressed in the group discussions with the GPTs. The most frequently reported obstacles were:

- It is not easy for patients to give adequate feedback
  ...Often, it just didn’t work...although when I asked a patient to be really honest in answering, I did get more variance in answers...

- Patients give socially desirable answers, possibly due to the fact that they know that it is not anonymous
  ...I know it was a bad consultation... but the patient was still positive about it, which was just hard to believe...

- Within a doctor-patient relationship it makes GPTs too vulnerable
  ...It’s not good for a doctor-patient relationship... you have to keep a professional distance...

- The majority of the relatively simple consultations did not provide ambiance for feedback
  ...Some of the consultations were so easy that feedback was not appropriate, and definitely not useful...

- Mixed thoughts about the intention of the programme
  ...I doubt whether the programme will benefit my communication skills...it will probably just provide material for the research group...

- Full agenda. Especially the recommended number of 30 questionnaires was considered to be too many
  ...All those questionnaires... it felt like a real burden in addition to all the other work...
Figure 2. Box-and-whisker plot of the number of questionnaires received by each GPT

Box shows the 25, 50 (median) and 75 cumulative relative frequencies (percentiles), whiskers represent the range.
Table 3. Evaluation of the programme by GPTs (n = 46)

<table>
<thead>
<tr>
<th>Open format questions</th>
<th>Distilled categories</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>What was your opinion about patient feedback at the beginning of the programme?</td>
<td>Positive / Very positive</td>
<td>82.0</td>
</tr>
<tr>
<td></td>
<td>With reservations</td>
<td>18.0</td>
</tr>
<tr>
<td>Do you consider patient feedback useful?</td>
<td>Yes</td>
<td>45.6</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>54.4</td>
</tr>
<tr>
<td>Were there any reasons for limited effect?</td>
<td>Positively biased answers</td>
<td>39.1</td>
</tr>
<tr>
<td></td>
<td>Patients not being able</td>
<td>26.1</td>
</tr>
<tr>
<td></td>
<td>No specific reason</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td>No limitations</td>
<td>21.7</td>
</tr>
<tr>
<td>How did it feel to receive PF (vulnerability)?</td>
<td>Vulnerable</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td>At ease</td>
<td>37.0</td>
</tr>
<tr>
<td></td>
<td>Not categorized</td>
<td>56.8</td>
</tr>
<tr>
<td>What obstacles did you encounter in organising PF?</td>
<td>No available room</td>
<td>8.7</td>
</tr>
<tr>
<td></td>
<td>Time constraints</td>
<td>30.4</td>
</tr>
<tr>
<td></td>
<td>Motivation of GPT</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td>Motivation of patients</td>
<td>8.7</td>
</tr>
<tr>
<td></td>
<td>Unspecified general</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>32.6</td>
</tr>
<tr>
<td>Did you complete the assignments and the related topic discussions with your GP?</td>
<td>Yes</td>
<td>37.0</td>
</tr>
<tr>
<td></td>
<td>Not all</td>
<td>32.6</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>30.4</td>
</tr>
<tr>
<td>Did you find these topic discussions useful?</td>
<td>Yes</td>
<td>47.6</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>52.4</td>
</tr>
<tr>
<td>Do you think the number of 30 questionnaires is adequate?</td>
<td>Yes</td>
<td>15.2</td>
</tr>
<tr>
<td></td>
<td>No (too many)</td>
<td>78.3</td>
</tr>
<tr>
<td></td>
<td>Not categorized</td>
<td>4.5</td>
</tr>
<tr>
<td>Did it interfere with your consulting hours?</td>
<td>Yes</td>
<td>26.5</td>
</tr>
<tr>
<td></td>
<td>A little</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>16.3</td>
</tr>
<tr>
<td></td>
<td>Not categorized</td>
<td>53.0</td>
</tr>
<tr>
<td>How much time did it take to organize and ask patients for feedback?</td>
<td>&lt; 5 minutes</td>
<td>17.9</td>
</tr>
<tr>
<td></td>
<td>5 minutes</td>
<td>53.6</td>
</tr>
<tr>
<td></td>
<td>5-10 minutes</td>
<td>28.6</td>
</tr>
</tbody>
</table>
Discussion and conclusion

Discussion
Both the exploratory and the feasibility study showed that patients and GPTs approve of the concept of patient feedback, and are willing to provide and receive patient feedback. The time-burden and organisational obstacles are thought to be of minor importance. Some consultations appeared to be less suitable for patient feedback, especially those that concerned relatively simple problems. These aspects were all taken into account in the development of the patient feedback training programme.

However, several obstacles were identified: patients provided very positive feedback which made it difficult for the GPTs to identify learning potentials, GPTs tended not to ask for feedback on difficult consultations, GPTs asked for feedback from a median of 18 instead of the planned 30 patients, and many GPTs considered the patient feedback questionnaires as a research tool rather than an educational tool. We will summarize how these obstacles could be overcome.

Feasibility of the modified PPPC Questionnaire
The GPTs indicated that the patients were generally positively biased in their comments, which may be a drawback with regard to learning potentials. It is well known that patients often give socially desirable answers, especially if they have to give their opinion about their own GP. Patients usually are also less critical than colleagues or trainers\textsuperscript{22}, and this has specific consequences for this training programme. If GPTs receive feedback by means of positively biased answers, it may not stimulate them to develop their consultation skills; in fact, it might even be counter-productive. Other authors have suggested that even small differences in the assessments made by patients, especially at the upper end of the scale, might have substantial implications\textsuperscript{24}. However, we should also take into consideration the possibility that patients might actually be truly satisfied with the consultation skills of the GPTs. On the other hand, despite this
general satisfaction, there may be levels of specific dissatisfaction that lie underneath\textsuperscript{25}.

\textit{Content-relatedness}
One of the more striking aspects was that, according to the GPTs, patient feedback is content-related. Apparently, when GPTs have to deal with relatively simple medical issues, patient feedback is not very appropriate or useful. This was repeatedly mentioned by both the GPTs and the patients. Although GPTs are generally well trained to investigate even the simple cases to look for hidden agendas, questions or anxieties, in daily practice the simple cases were not always appropriate for acquiring feedback. In view of the common belief that GPTs will take a high share in seeing patients with relatively simple medical problems, they should take care to select those patients from whom they would like to receive feedback.

\textit{Personal reluctance}
Only 6\% of the GPTs stated that they felt vulnerable in acquiring patient feedback (Table 3), although we suspect that this percentage might actually be much higher. Doctors receive feedback with a mixture of eagerness and anxiety. The fact that only 15\% of the questionnaires (instead of the suggested 50\%) were handed out to selected patients, implies that GPTs might find it difficult to ask patients for feedback. Patient feedback is meant for educational purposes, and the GPTs indicated that it is especially effective if it contains criticism. Positive feedback can also be beneficial, because it will reassure the GPT, and may stimulate the use of specific skills. Because the GPTs selected patients from whom they wanted to receive feedback, the aggregated scores on the questionnaires are, per definition, not suitable for assessment purposes or for benchmarking.

There is therefore a paradox that makes this programme interesting and complicated at the same time: on the one hand, the GPTs wanted to receive positive feedback with regard to their communication skills, but, on the other hand, they also indicated that it is especially critical feedback from which they can learn most.
Chapter 2

Training programme
In general, the GPTs and patients worked intensively on the assignments during the programme, and have demonstrated that it is feasible. Some GPTs were confused about the main objective of the programme. Because researchers, and not teachers, gave the training and did the interviewing, a number of GPTs felt that the programme was organised for the benefit of our research, and not so much as an aid to acquiring consultation skills.

Organisation
Because some GPTs tended to hand out the questionnaires to the relatively ‘easy’ patients, and some GPTs even struggled to hand out any questionnaires at all, our decision to let the GPTs decide which patient to select for asking for feedback is debatable. An alternative option would be to hand out the questionnaires to every patient and independently of the GPT (for example by a research assistant). Despite the obvious greater number of received questionnaires, we have rejected this option, because we expect the GPTs to yield more valuable feedback when they themselves select patients. In our experience, handing out questionnaires randomly or consecutively to all patients even raises the already existing ceiling effects. However, involving a research assistant or trained practice assistant may be a real alternative, especially when GPTs appear to struggle in completing their assignment. Another way in which to enhance the objectivity of the acquired feedback would be to make it more anonymous for the patients. They could then fill in the questionnaire at home and send it to the practice by post. However, this would hamper the comparison of patient and doctor opinions about a specific consultation, because of the delay in receiving the feedback. Another drawback of having the questionnaires posted to the practice would be the generally expected low response rate.

Conclusion
In conclusion, this study revealed that the concept of patient feedback is generally feasible for both patients and GPTs. However, major improvements have to be made in order to enhance the educational potential, and thereby the benefits.
Practice Implications

Based on the key learning points described above, we have made the following recommendations for improving the learning potential of the patient feedback programme and the questionnaires. The patients should be told explicitly that an honest reply would be appreciated, and might be beneficial for educational purposes. Moreover, the GPTs have to be prepared for the type of positively biased answers they can expect from patients. As a result, if a patient’s answer to one of the questions is not completely positive, the GPTs might already find an indication for substantial improvement here. The GPTs should hand out the questionnaires during the more challenging consultations, and they should also be trained to decide for themselves from which patient they expect to receive relevant feedback, i.e. those with complex problems, or to hand out a questionnaire when there is a feeling of GPT or patient dissatisfaction. This implies an additional advantage, in that we expect that the GPTs will need to hand out less questionnaires, which is in line with their wishes. Fully embedding patient feedback in the GP training programme means that GPTs and GP trainers should receive training and education from the teaching staff, and not from the researchers.

When we have integrated these learning aspects in our programme, we will plan a randomised controlled trial in which we will investigate whether patient feedback actually enhances the communication skills of GPTs. The relevance and objectivity of the patient feedback obtained from the questionnaires can then be compared with other objective measures, such as assessment of videotaped consultations.
References


22. Braend AM, Gran SF, Lindbaek M. [Patients--a useful resource when evaluating medical students' clinical practice?]. Tidsskr Nor Laegeforen 2006; 126(16):2122-2125.


CHAPTER 3

Validity aspects of the Patient Feedback questionnaire on Consultation skills (PFC), a promising learning instrument in medical education

Marcel E. Reinders
Annette H. Blankenstein
Dirk L. Knol
Henrica C.W. de Vet
Harm W.J. van Marwijk

Abstract

Objective
A focus on the communicator competency is considered to be an important requirement to help physicians to acquire consultation skills. A feedback questionnaire, in which patients assess consultation skills, might be a useful learning tool. An existing questionnaire on patient-centeredness (PPPC) was adapted to cover the ‘communicator’ items in the competency profile. We assessed the face and content validity, the construct validity and the internal consistency of this new Patient Feedback on Consultation skills (PFC) questionnaire.

Methods
We assessed the face validity of the PFC by interviewing patients and general practice trainees (GPTs) during the developmental process. The content validity was determined by experts (n=10). First-year GPTs (n=23) collected 222 PFCs, from which the data were used to assess the construct validity (factor-analysis), internal consistency, response rates and ceiling effects.

Results
The PFC adequately covers the corresponding ‘communicator’ competency (face and content validity). Factor analysis showed a one-dimensional construct. The internal consistency was high (Cronbach’s alpha 0.89). For the single items, the response rate varied from 89.2% to 100%; the maximum score (ceiling effect) varied from 45.5% to 89.2%.

Conclusion
The PFC appears to be a valid, internally consistent instrument.

Practice Implications
The PFC may be a valuable learning tool with which GPTs, other physicians and medical students can acquire feedback from patients regarding their consultation skills.
Introduction

Feedback plays an important role in the communication training programmes for general practice trainees (GPTs). While much of this feedback is provided by the educators, feedback from the patients themselves may be a useful perspective from which to assess the GPTs’ consultation performance. Getting to know what patients want or think about a visit to the doctor may benefit both the patient and the doctor\(^\text{1-3}\).

Several studies have been reported on patient satisfaction and accompanying questionnaires\(^\text{4-7}\). For our educational purposes, we needed a patient feedback instrument that, beyond the concept of mere patient satisfaction, focuses on the actual performance of a physician. Surveys have shown that patients are capable of giving this specific kind of feedback\(^\text{8-13}\).

Furthermore, we considered it of importance that the items contained in the questionnaire would reflect the content and meaning of the key competency ‘communicator’ (Table 1). Based on the CanMeds report (Canadian Medical Education Directives for Specialists, an initiative to make medical training programmes being responsive to societal needs)\(^\text{14}\) it is part of the new competency profile for GPs that has been developed recently by the Dutch general practitioners vocational training institutes, and serves as a framework for education and research programmes. We adapted and extended an existing questionnaire, the Patient Perception of Patient-Centeredness questionnaire (PPPC) that seemed to fit to some extent within our concept.

Recently published studies have validated questionnaires for the ratings of physicians’ communication skills by patients\(^\text{15,16}\). However, in these studies the patients and the doctors (North-American physicians of mixed boards including general practitioners) agreed on very high ratings of the consultation skills (ceiling effects). Since we will not only use the PFC for the assessment of consultation skills, but also as a learning tool that can be applied in our vocational training programme for GPTs, the ratings should not be too high, because there must be room for improvement. Moreover, we have learned from
an earlier study\textsuperscript{17} that GPTs are easily discouraged from continuing to ask patients for feedback when there is a lack of sufficient critical feedback. The challenge in this study was therefore to develop a questionnaire that adequately reflects the key competency ‘communicator’, that is void of ceiling effects as much as possible, and that is appropriate for use in our vocational training institute. In this paper we describe the development of this questionnaire, the ‘Patient Feedback on Consultation skills (PFC) questionnaire’, and present the results of a validation and reliability study.

\begin{table}[h]
\centering
\caption{Key competency ‘doctor-patient communication’}
\begin{tabular}{p{0.9\textwidth}}

A general practitioner:

\begin{itemize}
\item Creates effective therapeutic relationships with patients
  \begin{itemize}
  \item Creates trust
  \item Shows concern with patients (and families) and creates grounds for a long trusting therapeutic relationship
  \end{itemize}
\item Uses communication skills effectively
  \begin{itemize}
  \item Adapts language to the age, gender, ethnic and cultural background, and emotion of the patient
  \item Explains the structure of the consultation
  \end{itemize}
\item Encourages patients to actively participate in decision-making
  \begin{itemize}
  \item Explores the patient’s views by actively listening, and clarifies the reason for help
  \item Encourages the patient to respond to the questions asked, the information given and the diagnosis
  \item Advises the patient about possible treatment options and helps the patient to make choices
  \item Delivers and organizes information, and systematically checks that the information is well understood
  \item Discusses the practicality of the therapeutic plan
  \end{itemize}
\end{itemize}
\end{tabular}
\end{table}
Validity of the questionnaire

Methods

General design, setting and data-collection
In 2005, the patient feedback training programme for GPTs was introduced in the VU medical center in Amsterdam. Simultaneously, a study was carried out to assess its feasibility\(^\text{17}\). In 2006, 23 first-year GPTs completed this programme, in which they were trained how to acquire patient feedback by means of a questionnaire (the PFC). To enhance the learning potential, they were advised to ask those patients from whom they expected to get relevant and critical feedback, to fill in the PFC. If a patient was too young, the accompanying adult filled in the questionnaire. The patients completed the PFC at the practice, immediately after consulting the GPT, and returned it to the practice assistant in a sealed envelope. The GPT completed the self-assessment version of the PFC. At the end of the training programme, the GPTs returned all PFCs to the research team, and filled in an evaluation form concerning baseline characteristics and feasibility aspects.

Development of the PFC
Selection of questions
The development of the PFC was based on the following principles: its content would reflect on the recently developed competency profile for doctor-patient communication and would elicit sufficient critical feedback from patients to ensure the learning potential would be optimal.

The first draft version of the PFC included items from an existing questionnaire, the PPPC (9-item version, 4-point scale)\(^\text{18}\), which is a validated, single-scale instrument for the measurement of patient-centeredness\(^\text{19,20}\). It has two parallel versions: a patient feedback version and a physician self-assessment version. Several questions were added to ensure a complete coverage of the ‘communicator’ competency profile. Experts from within and outside our institute were asked to comment on the proposed draft version, and subsequent adjustments were made until there was agreement on the version to be used in the first preliminary study.
Refinement of the questionnaire

In a first preliminary study, representatives of the study population (51 patients and 17 GPTs) were interviewed in a semi-structured way, to investigate whether the questionnaire would provide the GPTs with useful feedback (face validity). They could indicate any difficulties they had in answering specific questions, and give suggestions for improvement. We also included comments from two representatives of the Board of the Dutch Patient Federation (NCPF). In group interviews, the GPTs were invited to comment on the questionnaire with regard to relevance and ambiguity, and this led to certain amendments to the items in the questionnaire. A previous feasibility study had already demonstrated that the patients took 3-4 minutes to complete the questionnaire, which they considered to be only a small burden.\(^17\)

From the questionnaires that had been filled in by the patients in a second preliminary study (48 GPTs, 878 patients), we were able to assess the response patterns and the non-response rate. If there was a strong ceiling effect or a high non-response rate for certain questions, they were omitted or modified. For example: for the question “Did you have confidence in this doctor?” there was an average satisfaction score of 100%, so it was therefore omitted. Another example: for the question “After visiting this doctor, are you able to cope with your illness?” the non-response rate was 50%, so this question was also omitted. The wording of the questions was found to influence the response pattern. For example: “Were you given the opportunity to ask questions?” was changed into “Did the doctor give you the opportunity to ask questions?”, which resulted in a maximum satisfaction score of 90% and 77%, respectively. In general, questions were phrased in an active way, so that the patients could comment on the actual performance of the doctor, instead of merely giving an impression of their satisfaction. Preliminary versions of the questionnaire, containing the questions that were omitted or modified, are available on request.

Table 2 presents the final version of the PFC. It is a self-administered questionnaire, with instructions and room for baseline patient characteristics on the cover. All 16 questions are listed on one page. The questionnaires are
numbered (and therefore not anonymous), in order to provide specific consultation-related feedback.

Table 2. The questionnaire for Patient Feedback on Consultation skills (PFC); 4-point scale (completely-mostly-a little-not at all)

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. To what extent was your main problem(s) discussed today?</td>
</tr>
<tr>
<td>2. How satisfied were you with the discussion of your problem?</td>
</tr>
<tr>
<td>To what extent did:</td>
</tr>
<tr>
<td>3. ... the doctor listen to what you had to say?</td>
</tr>
<tr>
<td>4. ... the doctor explain this problem to you?</td>
</tr>
<tr>
<td>5. ... you and the doctor discuss your respective roles?</td>
</tr>
<tr>
<td>6. ... the doctor explain treatment?</td>
</tr>
<tr>
<td>7. ... the doctor explore how manageable this treatment would be for you?</td>
</tr>
<tr>
<td>8. How well do you think your doctor understood you today?</td>
</tr>
<tr>
<td>9. To what extent did the doctor discuss personal or family issues that might affect your health?</td>
</tr>
<tr>
<td>10. Was there an atmosphere of trust during the consultation?</td>
</tr>
<tr>
<td>To what extent did:</td>
</tr>
<tr>
<td>11. ... the doctor show his/her concern?</td>
</tr>
<tr>
<td>12. ... the doctor invite you to ask all the questions you wanted to ask?</td>
</tr>
<tr>
<td>13. ... the doctor give you clear information and explanation?</td>
</tr>
<tr>
<td>14. ... the doctor act in a structured way?</td>
</tr>
<tr>
<td>15. ... the doctor give you new or better insight into your problem?</td>
</tr>
<tr>
<td>16. ... the doctor give you clear treatment advice?</td>
</tr>
</tbody>
</table>

Q 1-9 from: Stewart et al. *Patient Perception of Patient Centeredness*
Validity and internal consistency of the final version of the PFC

Content validity

Further confirmation of the content validity of the final version of the PFC was obtained by asking a panel of 10 experts from six University general practice departments (physicians, behaviour scientists and psychologists who are all involved in the vocational training of GPTs, and by the majority of whom were involved in the development of the Dutch competency profile for graduated GPs) whether the items in the PFC would sufficiently cover the content and meaning of the key competency ‘communicator’. These experts gave their comments independently, but the research group was not blinded for their answers.

Construct validity

We hypothesized beforehand that the two concepts, ‘patient-centeredness’ and the key competency ‘communicator’, would overlap, so that the PPPC and additional questions could be combined. We first performed an exploratory item factor-analysis on the matrix of polychoric correlations to determine whether the original single-scale construct of the PPPC was maintained, or that more factors could be extracted. Parameters were estimated with the weighted least squares, with the mean and variance correction method, and Promax rotation was performed to enhance the interpretation. The eigen-values were visualized in a scree plot; eigen-values above than 1.0 were considered to be relevant for assessment of the interpretability\(^ {21}\). We then performed a confirmative item factor analysis to assess how the PPPC (Q1-9) and the additional questions (Q10–16) would correlate, i.e. would they measure the same, or different constructs. Parameter estimates were obtained with the maximum likelihood method. Item factor analysis was performed with the Mplus 3.13 programme equipment \(^ {22}\).

Internal consistency

Internal consistency implies that the items measure the same underlying construct and that the sum-score of the items results in an interpretable score. We calculated the Cronbach’s alpha (which measures the extent to which item responses correlate with each other), and considered a value of 0.70 – 0.90 \(^ {23}\) to
be adequate. This was assessed for the PFC and compared with the translated PPC. Item-total correlations were calculated, and values above 0.3 were considered as evidence to support the reliability of the scale.

**Response rate and ceiling effects**

We analysed missing values (with the Mplus 3.13 programme) at item level and at patient level, because this gives an indication of the applicability of the questionnaire. We calculated the total percentage of missing values (total % of missings = sum of all missings / [222 questionnaires x 16 items]). The response rate was determined according to the proportion of unanswered questions per item. With regard to the assessment of ceiling effects on the single items, we present only the percentage of maximum scores in the text; the calculated percentages of all categories are presented in Table 3. The complexity of a consultation (according to the patients) was correlated with the percentage of PFCs with maximum scores on all items.

**Practical burden**

On the evaluation form the GPTs could indicate the amount of time they needed to explain the procedure and to recruit the patients.

**Table 3. Answering pattern and response rate for the PFC (in %), n=222**

<table>
<thead>
<tr>
<th>question</th>
<th>response</th>
<th>non-response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>completely</td>
<td>mostly</td>
</tr>
<tr>
<td>Q1</td>
<td>68.9</td>
<td>27.5</td>
</tr>
<tr>
<td>Q2</td>
<td>81.1</td>
<td>15.3</td>
</tr>
<tr>
<td>Q3</td>
<td>89.2</td>
<td>9.5</td>
</tr>
<tr>
<td>Q4</td>
<td>57.2</td>
<td>32.4</td>
</tr>
<tr>
<td>Q5</td>
<td>70.3</td>
<td>23.0</td>
</tr>
<tr>
<td>Q6</td>
<td>56.8</td>
<td>32.4</td>
</tr>
<tr>
<td>Q7</td>
<td>59.0</td>
<td>22.5</td>
</tr>
<tr>
<td>Q8</td>
<td>64.9</td>
<td>21.2</td>
</tr>
<tr>
<td>Q9</td>
<td>45.5</td>
<td>22.5</td>
</tr>
<tr>
<td>Q10</td>
<td>77.0</td>
<td>19.4</td>
</tr>
<tr>
<td>Q11</td>
<td>74.8</td>
<td>22.1</td>
</tr>
<tr>
<td>Q12</td>
<td>68.0</td>
<td>23.4</td>
</tr>
<tr>
<td>Q13</td>
<td>75.2</td>
<td>21.2</td>
</tr>
<tr>
<td>Q14</td>
<td>72.5</td>
<td>23.4</td>
</tr>
<tr>
<td>Q15</td>
<td>49.1</td>
<td>28.4</td>
</tr>
<tr>
<td>Q16</td>
<td>70.7</td>
<td>22.1</td>
</tr>
</tbody>
</table>
Chapter 3

Results

A total of 222 patients completed the final version of the PFC, 62.2% of whom were female, and the mean age was 41.1 years (SD 19.4). The 23 general practices were situated in a large city (36.9%), a small town (47.7%) or a rural area (15.3%). The complexity of the consultations, according to the patients, was: high (17.4%) medium (48.4%) or low (34.3%). In 49.8% of the consultations the patient visited the GPT for the first time.

Content validity
To obtain confirmation of the face and content validity of the PFC, we gathered a final round of expert opinions. According to their comments, the PFC correlated sufficiently with the key competency ‘communicator’, from which we concluded that no changes to the PFC were needed. A summary of the various (mild) comments are reported here. Relatively under-represented in the PFC was the reflection of the patient’s emotions or feelings; this was mentioned by two experts. Question 2 (satisfaction aspects), question 9 (personal or family issues) and question 15 (insight) were considered to be relatively redundant in covering the corresponding items of the key competency; this was mentioned by 4, 4, and 3 experts, respectively. Other remarks concerned the length of the questionnaire (‘too long’) and the wording of some of the questions (‘too difficult for patients’). These remarks were duly noted and will be taken into consideration in possible future developments.

Construct validity
In the exploratory item factor analysis we found one eigen-value of 8.90 and a second eigen-value of 1.21. These factors were not specifically represented in items of the PPPC or the additional questions. The scree plot (Figure 1) shows one single prominent factor and no other distinct break. The percentage of explained variance for the first factor was 55.64. The Promax rotation did not enhance the interpretation of the factors.
In the confirmative factor analysis, the correlation between the original construct (measured with the items of the translated PPPC [Q1-9]) and the construct (measured by the additional questions [Q10-16]) was 0.97. This means that the old and the new items measure the same construct.

**Internal consistency**
Cronbach’s alpha was 0.89 for the PFC and 0.83 for the translated PPPC. The item-total correlation ranged from 0.45 (question 11) to 0.67 (questions 9 and 13).

---

**Figure 1. Scree plot of eigen-values of the PFC**

![Scree plot of eigen-values of the PFC](image)
Response rate and ceiling effects

The total percentage of unanswered questions was 2.6. Table 3 shows that the percentage of unanswered questions (response rate) varied from 0.0 (question 1 'problem discussed' and question 3 'listening') to 10.8 (question 9 'personal or family issues'). Also shown in Table 3 is the response pattern for the PFC. The percentage of maximum scores ('completely'), or ceiling effect, varied from 45.5 (question 9 'personal or family issues') to 89.2 (question 3 'listening'). The mean percentage of maximum scores for single items was 67.5 (SD 11.6).

The percentage of patients who gave all the items on the questionnaire the maximum score on the scale (completely satisfied) was 17.1. Of these patients, 36.8% added very positive personal comments about the consultation skills of their GPT. The majority (66.7%) of the PFCs with a maximum score for all items concerned consultations that were considered by the patient and/or the GPT to be relatively 'easy'.

Practical burden

The average time needed for GPTs to inform and include a patient was 5–10 minutes. The GPTs repeatedly stated that an assignment to include 20 patients would be a considerable burden on their already busy time-schedule. The time patients needed to complete the questionnaire was 3-4 minutes (preliminary version of the PFC).
Discussion and conclusion

Discussion
This study focused on the validation of a patient feedback questionnaire, the PFC, as a feasible learning instrument for the development of consultation skills. It was aimed to focus on many aspects of the consultation skills and the subsequent phases of a consultation, and ultimately to help physicians to reflect on their consultation skills and guide discussions about these skills with their trainers or their patients.

The one-dimensional construct of the original PPPC (patient-centeredness) has been maintained in the PFC, as was indicated by factor analysis. We therefore consider the concept of patient-centered consultation to be interchangeable with the key competency ‘communicator’. Although we identified a second small factor that loaded for 1.21 on the PFC, we could not derive a sound theoretical construct for this factor. From the Cronbach’s alpha for the PFC we can conclude that its internal consistency remains high (0.89), which supports its reliability.

GPTs and patients contributed to the face validity and content validity during the development process. From a final round of expert opinions, we concluded that the content and meaning of the recently developed competency profile for GPs in the Netherlands (key competency ‘communicator’) is adequately covered by the PFC. Although some experts mentioned that certain questions (i.e. 2, 9 and 15) might be relatively redundant, we included all nine original PPPC items. A matter of concern is that the wording of the PFC might still be too complex for some patients.

Although we were satisfied with the generally high response rate for the single items, it appeared that question 9 (personal and family issues) had the highest non-response rate (10.8%), which implies that this question might not be applicable in all situations. For future use, the applicability might be enhanced if this question is changed into: ‘To what (necessary) extent did the doctor...?’.
Our results are in line with a Canadian study carried out by Campbell et al\textsuperscript{15}, in which a tool for the measurement of physician communication skills is presented. This supports the content validity of our mutual studies, although their data were derived from medical specialists and general practitioners in Canada, whereas our data were derived from GPTs in the Netherlands.

One of our goals in developing the PFC was to design an instrument that did not have strong ceiling effects caused by (overly) positive answers. This is a serious drawback that is commonly encountered in the use of questionnaires to assess patient satisfaction. For example, in the recently presented results of the study of Campbell et al\textsuperscript{15}, patients gave high ratings for consultation skills (item scores varied from 4.40 to 4.70 (scale 1-5). In another recent study, by Makoul et al\textsuperscript{16}, reported feedback ratings varying from 4.44 to 4.81 (scale 1–5). This is an interesting, but largely underdeveloped area of research in itself: what constitutes more effective learning material for GPTs: positive or negative feedback? In earlier studies we noticed that unrealistically positive feedback demotivated the GPTs from continuing to hand out questionnaires. In the present study we have tried to avoid this phenomenon, but have not succeeded entirely. The mean percentage of items on the PFC with a ‘completely satisfied’ score was 67.5, which we consider to be acceptable.

The ceiling effects might be caused by the so-called acquiescent response set, i.e. a sub-set of patients who automatically fill in a questionnaire in just one column (with a maximum score), and also by a sub-set of patients who give only socially desirable maximum scores. We cannot distinguish between these options and genuinely positive opinions, but a substantial number of the PFCs with maximum scores only, contained very positive additional written comments from the patients (36.8%). If we subtract these PFCs from the total percentage of PFCs with maximum scores (17.1), we could roughly estimate the acquiescent response set and/or social desirability to be 10%, which is a serious obstacle.

Of course, patients might be truly satisfied with the communicative aspects of a consultation with a physician. However, from the perspective of the educators that regularly assess video-taped consultations, there may still be considerable room for improvement in many cases. This also could imply that
feedback from patients who are not ‘completely satisfied’ but even ‘mainly satisfied’ can be interpreted by GPTs as a serious opportunity for improvement and learning. Makoul et al favour the idea of dichotomising the outcome into ‘excellent’ ratings (maximum scores) versus all other ratings, because of the highly skewed outcome (ceiling effects)\(^\text{16}\). We find this idea appealing, and expect that for the GPT a dichotomised outcome is more educative.

Despite the demonstrated validity of the PFC and the achieved reduction of ceiling effects within the refinement process, this study has certain limitations. Although in the developmental process of the PFC the patient’s involvement was warranted, the PFC is still a questionnaire that has been designed by doctors. Therefore, there might be a difference between the patient’s expectations and the doctor’s preferences (competency profile).

Furthermore, because patients were included by the GPTs themselves (to enhance the learning potential), this also implies that in our patient feedback training programme the PFC does not provide an objective assessment of the GPT’s consultation skills.

The design of consultation-specific patient feedback limits the anonymity of the patients; the advantages (enhanced learning potential, high response) and disadvantages (socially desirable answers) of which have been addressed in an earlier paper\(^\text{17}\).

Another limitation of the study might be that we did not collect specific information about the context of the consultations. Diagnostic uncertainty might impair the demonstration of a GPT’s consultation skills\(^\text{24}\).

Because less complicated consultations seem to be related to more PFCs with a maximum score on all items, we need to convince the GPTs who participate in our patient feedback training programme to select the more challenging consultations and to identify those patients who can make a real contribution.

**Conclusion**

The results of this study have confirmed that the PFC is a single-scale instrument that is both valid and internally consistent. Further validation elsewhere is needed to increase its external validity. It has the potential to
Contribute to the acquisition of communication skills by GPTs, it is applicable in most consultations, has acceptable ceiling effects, and is only a small burden to complete. There is, however, a continuing challenge to acquire feedback from patients that is relevant and critical, because it is that kind of feedback that is likely to provide the GPTs with a rich learning potential.

Practice implications
We will continue to make use of the PFC in the patient feedback training programme for GPTs, and it will be a fixed topic within the vocational training for GPTs. Although we have applied the PFC to GPTs and patients in general practice, the questions are certainly not limited to use in general practice education only. Medical School Boards and the residents training programmes for different medical specialties might also benefit, and because the PFC is available in Dutch as well as in English, this opens up possibilities for the application on a much wider (international) scale as well.
Validity of the questionnaire

References


Patient assessments of patient-centered consultation skills of general practice trainees

Marcel E. Reinders
Harm W.J. van Marwijk
Annette H. Blankenstein
Moira A. Stewart
Dirk L. Knol
Henriëtte E. van der Horst

Submitted
Abstract

Introduction
Teaching patient-centered consultation skills is high on the agenda of GP vocational institutes. We investigated the patients’ perspective on components of patient-centeredness: (a) ‘exploring the disease and illness experience’, (b) ‘finding common ground’ and (c) ‘personal context’.

Methods
A cross-sectional study based on 888 patient assessments. Forty-eight first-year general practice registrars (GPTs), at the VU medical center in Amsterdam, asked their patients to complete the Patient Perception of Patient-Centeredness questionnaire (item-score range: 1-4). The GPTs also completed a self-assessment version of the questionnaire. We applied regression analyses to identify variables that had effect on the patient assessments.

Results
The mean patient item-scores were 3.63 (SD 0.38). They scored high on component (a), but not so high on components (b) and (c). The mean GPT scores were 2.94 (SD 0.49), p<0.0001. Younger patients, male patients, lower language proficiency of the GPTs, and greater consultation complexity, were related to lower patient scores.

Conclusion
‘Finding common ground’ and ‘personal context’ are areas in which there is a need for change. Studying patient assessments of GPT consultation skills might form a basis for new learning strategies for GPTs, and might have implications for patient-doctor communication programmes.
Patient view on patient-centeredness

Introduction

In primary care, a patient-centered approach has been advocated as the preferred method of communication, because it is related to improved patient outcomes, such as satisfaction, enablement, compliance, better health status, less symptom burden and less diagnostic testing expenditures\(^1^\text{-}^4\). The patients themselves have a strong preference for a patient-centered approach\(^5^\text{-}^6\).

There is a growing tendency to include the patients’ views on general practice care and the quality of consultations in the training of young professionals\(^7^\text{-}^{14}\), creating opportunities for general practice trainees (GPTs) to learn from their patients’ feedback on their (patient-centered) consultation performance.

From the results of observational studies (interviews) there is evidence of what patients consider to be important components of patient-centeredness, and distinct differences are found between the perspectives of the patients and the perspectives of the doctors. The receptiveness of a doctor and the affective components of patient centeredness have been considered to be more important by GPs than by patients, but the patients showed greater endorsement of information\(^15\). However, less is known about the patients’ perceptions of these components of patient-centeredness within a setting of vocational training for general practice.

The acquisition of patient-centered consultation skills starts with learning about the patient’s evaluation of the patient-centeredness of the consultation. Because there is evidence that GP(T)s are better in finding a diagnosis than in giving information or shared decision-making\(^16\), we studied components of patient-centeredness, ‘exploring the disease and illness experience’, ‘finding common ground’ and ‘personal context’, from the patients’ perspective. Learning about the relative importance of these components might change the emphasis on certain aspects of the doctor-patient communication training programme.

GPTs at the VU medical center in the Netherlands are stimulated to acquire patient-centered consultation skills\(^17\), by means of validated feedback
questionnaires (Patient Perception of Patient-Centeredness [PPPC]) evaluating their consultation skills directly after a consultation.

In this study, we analysed the results of these questionnaires with regard to the components of patient-centeredness of the GPT consultations. We also analysed whether GPT characteristics, patient characteristics, and consultation complexity correlated with patient feedback scores, because this information could demonstrate that the evaluation of patient-centeredness might vary among patients, or might be consultation-specific. Doctor-related variables, for example, such as female gender, a younger age and better language proficiency, have been found to be correlated with better communication skills\(^1\). Therefore, we addressed the following research questions:

- What are the patient feedback scores on the PPPC, with regard to the components of patient-centeredness?
- Which variables are correlated with the patient assessments on the PPPC?

### Methods

**Design**

A cross-sectional study to investigate patients’ assessment of the components of patient-centered consultation skills. The analysis was based on the results of a total of 888 feedback questionnaires completed by patients.

**Setting**

From September 2005 until February 2006, 48 out of a total of 58 first-year GPTs at the VU medical center in Amsterdam, the Netherlands, attended a patient feedback training programme (10 did not attend for various reasons: pregnancy leave, different educational trajectories, or not completing the residency training). The patient feedback training programme provides the GPTs with theoretical and practical support in acquiring patient feedback on their patient-centered consultation skills\(^1\). During a period of three months, the GPTs
were instructed to ask (personally or intermediated by others, e.g. practice assistants) 30 patients to give feedback on their consultation skills, by completing the PPPC directly after the consultation, and returning the questionnaire in a sealed envelope to the practice assistant.

**Patients**

The GPTs could include patients from whom they wanted to receive feedback (which to our ideas enhanced the learning potential). Therefore, the selection of patients was not a random sample. As an indication, the GPTs were instructed to ask 30 patients to complete the PPPC (see assessment instruments). Insufficient command of the Dutch language was the only criterion for the exclusion of patients. If patients were too young, the accompanying adult could fill in the questionnaire. The patients gave oral informed consent to participate, and completed the PPPC in the practice. At the end of the training programme all the questionnaires were collected for analysis.

**Assessment instrument**

The questionnaire used in this study, the PPPC, was derived from the patient-centered clinical method developed by Stewart et al\textsuperscript{19}. The PPPC measures aspects of patient-centered communication within a consultation, and has been tested for validity and reliability. Patients can indicate their satisfaction concerning nine items on a 4-point scale (4=completely, 3=mostly, 2=a little, 1=not at all). The questionnaire has two parallel versions, one feedback version for the patient and one self-assessment version for the doctor.

The items in the PPPC refer to three components of the concept of patient-centeredness (see also the caption of Figure 1):

1. exploring both the disease and the illness experience (five questions: 1-4, 8),
2. finding common ground (three questions: 5-7),
3. personal context (understanding the whole person) (one question: 9).

The patient feedback version contains a cover page with questions on patient demographics, whether it was a first visit or a follow-up consultation, and the patient’s opinion about the complexity of the problem.
The self-assessment version had an attachment containing questions about the practical application of the questionnaire, such as: ‘Who recruited the patient?’.

**Analysis**
To evaluate patient feedback scores on the PPPC, we counted them per component of patient-centeredness. The patient scores and the GPT self-assessment scores are presented as a mean item-score (addition of single item-scores, range maximum=4 to minimum=1, divided by the number of questions involved).

For calculating the combined effect on the mean patient assessment scores (and on scores for the single components of patient-centeredness) the following variables were taken into account: patient characteristics and demographics (age, gender, location of practice), GPT characteristics (age, gender, work experience as a clinician, Dutch as first language [language proficiency]), motivational aspects (number of completed PPPCs they received from their patients), and consultation characteristics (who asked the patient, was it a first visit or a follow-up consultation, and how complex was the consultation according to the patients). We entered all variables with a value of p<0.2 into a backward elimination regression model.

**Results**

**Baseline characteristics**
The mean age of the patients was 39.8 years (SD 19.5), and the female/male ratio was 1.65. In 50.3% of the consultations the patient had visited the GPT for the first time. The practices were located in a large town (28.4%), a small town (60.0%), or in a rural area (11.6%). According to the patients, the consultations were highly complex (15.4%), of medium complexity (47.2%) or of low complexity (37.4%).

Most of the GPTs were in their early thirties (mean age 32.3 years [SD 5.7]), most of them were female (71.3%), and they had a mean work
experience of 3.5 years (SD 3.3). The mean number of PPPCs collected per GPT was 18 (SD 8.6, range 0–32).

**Assessments of the components of patient centeredness**

The calculated mean item-score for patient assessments was 3.63 (SD 0.38). The highest score was related to ‘exploring both the disease and the illness experience’ (3.78 [SD 0.34]), followed by ‘finding common ground’ (3.54 [SD 0.54]), and the lowest score was related to ‘personal context’ (3.17 [SD 0.96]).

The GPT self-assessment scores were much lower: the mean item-score was 2.94 (SD 0.49), ranging from 3.22 (SD 0.48), to 2.63 (SD 0.74), to 2.43 (SD 0.96), respectively (p<0.0001, independent-samples t-test).

Figure 1 shows pair-wise the patient assessment scores and the GPT self-assessment scores for the single questions on the PPPC. Because of the relatively high patient ratings, the scores were dichotomized into the maximum category ‘completely satisfied’ versus the other three categories. Figure 2 presents dichotomized pairs of patient scores and GPT scores for the three components of patient-centeredness. We found that the patient scores, although significantly higher, followed the same pattern as the GPT scores.

**Variables correlating with patient assessments**

Table 1 presents the mean item-scores of the patient assessments, for the several variables that were taken into account. Table 2 presents a linear regression model with patient scores as a dependent value. The age and gender of the patient, language proficiency of the GPT, and the complexity of the consultation, all had significant regression coefficients, and therefore were correlated with the patient scores. Itemizing the dependent value into the three components of patient-centeredness did not identify any other variables with a significant regression coefficient.
**Figure 1. Patient assessment scores and GPT self-assessment scores for the single items of the PPPC**

Scores are dichotomized: the maximum category (‘completely’), which is presented here, versus the other 3 categories (‘mostly – a little- not at all’).

Questions on the PPPC:
1. To what extent was your main problem(s) discussed today?
2. How satisfied were you with the discussion of your problem?
3. To what extent did the doctor listen to what you had to say?
4. To what extent did the doctor explain this problem to you?
5. To what extent did you and the doctor discuss your respective roles?
6. To what extent did the doctor explain treatment?
7. To what extent did the doctor explore how manageable this (treatment) would be for you?
8. How well do you think your doctor understood you today?
9. To what extent did the doctor discuss personal or family issues that might affect your health?
Figure 2. Patient assessment scores and GPT self-assessment scores for the three components of patient-centeredness

*error-bars represent standard deviation.
Table 1. Patient assessment scores

<table>
<thead>
<tr>
<th>category</th>
<th>Variables</th>
<th>range</th>
<th>n</th>
<th>patient assessment of PPPC&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Mean, SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
<td>Age&lt;sup&gt;2&lt;/sup&gt;</td>
<td>16-49</td>
<td>491</td>
<td></td>
<td>3.61 (0.34)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥50</td>
<td>290</td>
<td></td>
<td>3.67 (0.39)</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>male</td>
<td>314</td>
<td></td>
<td>3.59 (0.40)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>female</td>
<td>516</td>
<td></td>
<td>3.67 (0.36)</td>
</tr>
<tr>
<td></td>
<td>Practice location</td>
<td>city</td>
<td>242</td>
<td></td>
<td>3.59 (0.40)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>town</td>
<td>508</td>
<td></td>
<td>3.66 (0.37)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>village</td>
<td>101</td>
<td></td>
<td>3.60 (0.36)</td>
</tr>
<tr>
<td>GPT</td>
<td>Age</td>
<td>&lt;30</td>
<td>330</td>
<td></td>
<td>3.65 (0.32)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥30</td>
<td>487</td>
<td></td>
<td>3.63 (0.39)</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>male</td>
<td>202</td>
<td></td>
<td>3.63 (0.36)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>female</td>
<td>676</td>
<td></td>
<td>3.64 (0.39)</td>
</tr>
<tr>
<td></td>
<td>Work experience as a clinically</td>
<td>≤1</td>
<td>287</td>
<td></td>
<td>3.63 (0.34)</td>
</tr>
<tr>
<td></td>
<td>involved doctor (years)</td>
<td>1-5</td>
<td>268</td>
<td></td>
<td>3.69 (0.31)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥5</td>
<td>261</td>
<td></td>
<td>3.60 (0.43)</td>
</tr>
<tr>
<td></td>
<td>Language proficiency (Dutch as</td>
<td>yes</td>
<td>715</td>
<td></td>
<td>3.65 (0.34)</td>
</tr>
<tr>
<td></td>
<td>first language)</td>
<td>no</td>
<td>163</td>
<td></td>
<td>3.56 (0.48)</td>
</tr>
<tr>
<td></td>
<td>No. of received PPPCs completed by</td>
<td>≤10</td>
<td>65</td>
<td></td>
<td>3.72 (0.38)</td>
</tr>
<tr>
<td></td>
<td>patients</td>
<td>10-20</td>
<td>239</td>
<td></td>
<td>3.62 (0.40)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥20</td>
<td>579</td>
<td></td>
<td>3.63 (0.38)</td>
</tr>
<tr>
<td>Patient</td>
<td>Who included the patient?</td>
<td>GPT</td>
<td>726</td>
<td></td>
<td>3.64 (0.38)</td>
</tr>
<tr>
<td>selection</td>
<td></td>
<td>assistant</td>
<td>122</td>
<td></td>
<td>3.63 (0.40)</td>
</tr>
<tr>
<td></td>
<td>First consultation or follow-up</td>
<td>first</td>
<td>441</td>
<td></td>
<td>3.64 (0.37)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>follow-up</td>
<td>436</td>
<td></td>
<td>3.63 (0.38)</td>
</tr>
<tr>
<td></td>
<td>Complexity of the consultation</td>
<td>complicated</td>
<td>116</td>
<td></td>
<td>3.56 (0.40)</td>
</tr>
<tr>
<td></td>
<td>(patient report)</td>
<td>moderate</td>
<td>356</td>
<td></td>
<td>3.61 (0.40)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>simple</td>
<td>282</td>
<td></td>
<td>3.69 (0.31)</td>
</tr>
</tbody>
</table>

<sup>1</sup>Patient Perception of Patient-Centeredness; <sup>2</sup>patients younger than 16 were not included because a companying adult completed the questionnaire.
Table 2: Linear regression model of variables correlating with patient assessment scores

<table>
<thead>
<tr>
<th>variables</th>
<th>B</th>
<th>SE</th>
<th>p</th>
<th>confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>patient’s age*</td>
<td>-.813</td>
<td>.403</td>
<td>.044</td>
<td>-1.604 - -0.023</td>
</tr>
<tr>
<td>patient’s gender†</td>
<td>-.825</td>
<td>.251</td>
<td>.001</td>
<td>-1.317 - -0.333</td>
</tr>
<tr>
<td>GPT’s language proficiency‡</td>
<td>1.007</td>
<td>.316</td>
<td>.002</td>
<td>0.386 - 1.628</td>
</tr>
<tr>
<td>consultation complexity§</td>
<td>-1.200</td>
<td>.365</td>
<td>.001</td>
<td>-1.917 - -0.484</td>
</tr>
</tbody>
</table>

* age category 16-49 versus age category ≥50 (years), † male versus female, ‡ Dutch as a first language (yes versus no), § complex and moderate complexity versus relatively easy consultations.

Discussion

Main results
Learning the patients’ view with regard to the components of patient-centeredness is an important basis for acquiring patient-centered consultation skills. We investigated the patient assessments of the three components of patient-centeredness included in the PPPC. The scores for items concerning the components ‘personal context’ and ‘common ground’ were relatively low (for both patients and GPTs). The patient scores were significantly higher than the GPT self-assessment scores, in other words, the GPTs were more critical about the patient-centeredness of their consultation skills than the patients, which is as we expected. In general, however, the assessments followed the same pattern.

Interpretation of the results
The lower scores for items referring to the explanation and the manageability of the proposed treatment (‘common ground’), and the lower scores for the item referring to the personal and family issues (‘personal context’) might be caused by the tendency of physicians to focus on the medical content of a consultation.
This finding conforms with the results of earlier research\textsuperscript{16,20,21}, and with our personal experience as teachers, i.e. that first-year GPTs are preoccupied with the first phase of a consultation (finding a diagnosis), and that there is yet no room for finding agreement or giving information. For the patients, however, receiving information is important\textsuperscript{15,22}. It is likely that for comparable reasons personal and family issues were often not discussed, and therefore rated as low by the patients. It seems that GPTs consider this item to be context-related, i.e. less relevant when dealing with relatively simple medical issues (personal interviews with GPTs). In contrast to the results of other studies in which doctors clearly over-rated the amount of time and attention they spent on providing information\textsuperscript{23,24}, the GPTs in our study seemed to be well aware of their relative deficiencies.

**Correlations**
Younger patients and male patients appeared to be more critical in their assessments. However, although these differences in outcome were significant, they appeared to be quite small. It is reassuring that in general, patients seem to be able to give feedback on consultation skills, and there is no reason for GPTs to exclude patients for asking for feedback. GPTs who were lacking in language proficiency (whose first language was not Dutch) also had lower patient feedback scores. Other GPT characteristics, such as age, gender or work experience, were not related to patient feedback scores, in contrast to what has been reported in other studies\textsuperscript{18,25}. Furthermore, the complexity of the consultation was significantly associated with lower patient assessment scores. This indicates that patient-centered consultation skills are context-related, probably because there is a greater demand for skills, and therefore more opportunities for critical feedback\textsuperscript{26}.

**Limitations**
In this study, the patient scores for items related to ‘exploring the disease and the illness experience’ were relatively high (ceiling effect), which might limit the validity of the results. This finding is similar to the results of most previous studies\textsuperscript{11,27,28}. Although it could simply be suggested that the patient-centered
consultation skills appeared to be good, there is evidence from other sources (such as the training staff) that there is still room for improvement. The GPTs themselves think that these high scores do not reflect the reality of a consultation\textsuperscript{17}. Another limitation might be that, despite the large number of patients, the selection of patients is not a random sample. In order to enhance the educational aspect of feedback from patients, the GPTs were encouraged to include patients from whom they thought they would receive relevant feedback.

**Practice implications and future studies**

The patient assessments have indicated that the GPTs performed less well on the components ‘personal context’ and ‘common ground’. This may indicate that more emphasis should be laid on these components in the doctor-patient communication training programme. Studying the effect of this intervention on the skills of GPTs would be an interesting subject for future research. For educational purposes, however, as the GPT training advances, the focus should be on increasingly complex interactions, in keeping with the progress of the GPT. It will be interesting to investigate whether or not there would be a shift in the assessment of patient-centeredness if more experienced GPTs or registered general practitioners were the subjects of study.

**Conclusion**

This study focused on the patients’ view with regard to the components of patient-centeredness, within the GP vocational training setting. These aspects were scored relatively low by the patients. These findings may form an important basis for GPTs in acquiring patient-centered consultation skills, and also have practical implications for the design and content of patient-doctor communication programmes.
References


CHAPTER 5

Does patient feedback improve the consultation skills of general practice trainees? A controlled trial

Marcel E. Reinders
Annette H. Blankenstein
Henriëtte E. van der Horst
Dirk L. Knol
Piet L. Schoonheim
Harm W.J. van Marwijk

Medical Education 2010; 44: 156-164
Abstract

Introduction
This study aimed to assess if an additional patient feedback training programme leads to better consultation skills in General Practice Trainees (GPTs) than regular communication skills training, and whether process measurements (intensity of participation in the programme) predict the effect of the intervention.

Methods
We carried out a controlled trial in which two sub-cohorts of GPTs were allocated to an intervention group (n=23) or a control group (n=30), respectively. In 2006, allocated first-year GPTs in the VU medical center, attended a patient feedback training programme in addition to the regular communication skills training. The control group attended only regular communication skills training. Trainees were assessed by simulated patients who visited the practices and video-taped the consultations at baseline and after three months. The videotapes were randomly assigned to eight trained staff-members. The MAAS-Global instrument (range 0-6) was used to assess (a change in) trainee consultation skills. Results were analysed using a multilevel, linear mixed model analysis.

Results
Data on 50 GPTs were available for the follow-up analysis. Both the intervention group and the control group GPTs improved their consultation skills: the mean MAAS-Global scores for all participants were 3.29 (standard deviation [SD] 0.75) at baseline and 3.54 (SD 0.66) at follow up (p=.047). The improvement in MAAS-Global scores in the intervention group did not differ significantly from the improvement in the control group. The analysis showed a trend for intensity of participation in the patient feedback programme to predict greater improvement in MAAS-Global scores.
**Discussion**

Although the baseline scores were already in the high range, consultation skills in both groups improved significantly. This is reassuring for current teaching methods. The patient feedback programme did not improve consultation skills more than regular communication training. However, a sub-group of GPTs who participated intensively in the programme did improve their consultation skills further in comparison with the less motivated sub-group.
Introduction

That it is important that medical professionals acquire adequate communication skills is widely accepted because these skills are related to several major outcomes, such as enhanced patient satisfaction, therapy compliance and symptom relief, as well as a reduction in litigations and costs\(^1,2\). In the Netherlands some doubts have been raised about the effectiveness of conventional skills teaching because general practice trainees (GPTs) have been shown to demonstrate no improvement in communication skills during their training\(^3\).

A focus on the new communication competency\(^4\) and opportunities for self-directed learning\(^5\) in a rich, practice-based learning environment, are now considered to be essential requirements for education programmes focusing on communication skills. What is still unclear, however, is how each method contributes to this new learning trajectory and which tools help GPTs to improve their consultation skills.

A promising educational method that has received increasing attention in recent years is feedback from patients on the consultation skills of practising doctors and GPTs. Receiving (and giving) feedback and performance assessment are known to be strong motivators in the learning process\(^6,7\). GPTs frequently receive feedback from their general practitioner (GP) trainers or tutors, but the patients who consult the GPT are theoretically in the best position to give feedback. It has been recommended that patient feedback as an educational method should be included in the regular communication skills training programme\(^3,8,9\).

The methods applied in patient feedback interventions vary widely. The most common method to have been proved feasible consists of providing patient feedback on consultation skills obtained by means of a questionnaire\(^10-14\). Other methods include the presentation of individualised structured feedback reports\(^8,9,15\) and small group-meetings in which the results of feedback are discussed\(^16\). However, there is no clear evidence that any one of these approaches is superior to the others\(^17\).
From an earlier study focusing on the feasibility of a patient feedback programme, we concluded that the educational value was correlated to the context of the consultation\textsuperscript{18}. Consultation-specific feedback, provided by the patient shortly after the consultation, was considered to be effective by both the participants and the researchers.

Therefore, we added a patient feedback training programme to our doctor-patient communication skills education. In this programme, GPTs were trained to select and include patients from whom they believed they might learn the most. The instrument they used for this purpose was a validated questionnaire, the Patient Feedback Questionnaire on Consultation Skills (PFC)\textsuperscript{19}, the content of which focuses on general aspects of doctor-patient communication. To enhance the educational impact, the GPTs also completed a self-assessment version of the PFC and compared the two.

Although some studies have reported on the level of appreciation of patient feedback programmes or on changes intended or initiated as a result of patient feedback\textsuperscript{9,18,20}, there is still only limited and equivocal evidence of the actual effect of patient feedback interventions on the acquisition or improvement of consultation skills\textsuperscript{12,15,21-24}. Conflicting findings have shown that patient feedback delivered by means of individualised structured feedback reports did not improve the consultation skills of GPs\textsuperscript{15}, but did improve the consultation skills of GPTs to a certain extent\textsuperscript{23}.

Therefore, we carried out a controlled trial in which we studied the effect of an intervention (patient feedback training programme) in comparison with regular communication skills education (control group). We addressed the following research questions:

- To what extent will GPTs improve their consultation skills within the study period?
- Does an additional patient feedback training programme achieve greater improvement in the consultation skills of GPTs than regular communication skills training alone?
- What determinants (baseline characteristics, intensity of participation in the programme) predict the effect of the intervention, if any?
Chapter 5

Methods

Participants
All first-year GPTs who started their three-year postgraduate training in general practice in March 2006 at the VU medical center in Amsterdam were eligible for inclusion in this trial. Trainees were excluded if they had given up the vocational training course permanently or temporarily before the baseline-assessments, or if they had already attended a previous patient feedback training programme.

Allocation procedure
The regular educational training of GPTs is conducted by class (five classes, with seven to 13 GPTs in each). In order to prevent contamination, the randomisation of GPTs to the intervention group or the regular training (control) group could only be performed by class. Furthermore, classes with education programmes on the same day of the week were combined. The cohort was therefore dichotomised. An independent, blinded co-worker performed the allocation of GPTs into the intervention or control group.

Intervention group
During September - December 2006 the GPTs attended a patient feedback training programme which had been developed in 2005 at our institute\(^\text{18}\). The objective was to teach GPTs how to acquire relevant patient feedback on their consultation skills in daily practice by means of a procedure that was based on a validated questionnaire, the PFC\(^\text{19}\). The PFC contains 16 questions related to the doctor’s communication behaviour, scored on a 4-point scale. It is designed to cover all items of the competency profile ‘communicator’ (single construct) and has been found to be sufficiently reliable (Cronbach’s alpha 0.89). There are two versions, a patient feedback version and a doctor self-assessment version, across which questions are paired such as: ‘To what extent was (or were) your main problem(s) discussed today?’, and: ‘To what extent was (or were) your patient’s main problem(s) discussed today?’
The programme was delivered in addition to regular doctor-patient communication skills training and was supervised by the educational staff. Before the course started, staff members received instructions (delivered in a two-hour meeting), a protocol and a checklist of the items to be discussed with the participants.

The intervention started with a half-day instruction meeting. All course materials were provided in digital and print form. Using simulated patients (SPs), GPTs practised how to ask a patient for feedback and discussed how to organise patient feedback during daily practice. Over a subsequent period of three months the GPTs were required to ask for feedback from 20 patients from whom they expected to obtain relevant comments. Patients completed the PFC in the practice and returned it (in a sealed envelope) to the practice assistant. After comparing the pairs of feedback questionnaires and self-assessment questionnaires, the GPTs formulated learning points which they discussed with the GP trainers in their respective practices (and which were noted on the PFCs). At the end of the programme the GPTs returned all PFCs to the research team.

Control group
The control group did not attend the patient feedback programme, but undertook only the regular doctor-patient communication skills training. The latter consists of practice-based, experiential learning under the supervision of a GP trainer, who gives feedback on (video-taped) consultations of the GPTs’ real patients. For one day per week the GPTs attend group training courses at our institute, during which approximately 25% (two hours) of the time is spent on developing consultation skills. The subsequent phases of a consultation, interviewing techniques (summarising, reflecting, structuring) and a patient-centered approach, are studied and practised in real patient consultations and in simulated consultations.

Data-collection
Three SPs (standardised patients played by professional actors) were trained to enact six consultation scenarios of moderate and comparable complexity. They visited the 53 practices during the GPTs’ consultation hours. Visits were
announced and the GPTs gave their consent before participating. Only after the follow-up visits were the SPs allowed to give feedback to the GPTs. All consultations (one scenario at each visit) were videotaped. The SPs were assigned to the GPT practices in accordance with their own agendas, the GPTs’ agendas and the geographical locations of practices. They did not know which group the GPTs were allocated to. The three scenarios they played were performed ad random.

Assessments
Eight members of the education staff at our institute (five behavioural scientists, three GPs) assessed the videotaped consultations with the SPs. In order to enhance the quality of the measurements, the assessors were trained by an expert in the use of the MAAS-Global instrument in June 2007. They were required to study the manual and to score standardised videotaped consultations. During an instruction meeting, all assessors met the expert and discussed their findings.

After all follow-up assessments had been collected, we used a computerised random-number generator to assign the videotaped consultations (baseline and follow-up) to the eight assessors. All consultations were scored twice, independently, by two different assessors, who were blinded to each others’ scores and to the allocation of GPTs to the intervention or control groups.

Primary outcome measure (MAAS-Global score)
The primary outcome was the assessment of consultation skills with the MAAS-Global instrument, which was developed and validated at Maastricht University as a tool for educational purposes. It consists of 11 case-independent items, referring to communication skills, both in the specific phases of the consultation and in the entire consultation. Items are scored on a 7-point scale of 0-6 where 0=absent and 6=excellent. The end score is calculated as the sum-score of the single items, divided by 11.
Effectiveness (controlled trial)

Process measurements
We assessed the intensity of GPT participation in the programme (process measurements) based on: the number of PFCs that the GPTs collected; the number of learning points he or she formulated and whether the GPT discussed the learning topics with his or her GP trainer. Dichotomised scores (‘Z-scores’) were assigned for these three items and then added together (maximum score =3, minimum score=0).

Other determinants
The GPTs completed an evaluative questionnaire at the end of the study period. This contained questions about their baseline characteristics (age, gender, working experience as a doctor, practice location area, Dutch as first language).

They also completed the National Knowledge Test in General Practice Medicine, which was developed by the Foundation of United University Departments of General Practice (SVUH) to monitor the progress of GPTs in acquiring knowledge skills. A score that is less than the mean -1 standard deviation (SD) score of GPTs at a similar stage is considered to represent a failure (assessment in October 2006).

Statistical methods
We used a multi-level, linear mixed-model analysis with fixed effects (study group, intervention, GPTs) to study individual changes in overall MAAS-Global scores between baseline assessments and post-intervention assessments (p-values<.05 were considered to be significant). An additional advantage of mixed-model analysis of longitudinal data is that all measurements are used. In another multi-level linear regression model we studied the effect of process measures (intensity of participation in the intervention) on MAAS-Global scores over time.

The effect size \(d\) was measured by calculating the difference between baseline and follow-up MAAS-Global scores, divided by the pooled initial SD: a \(d\) of 0.2 was considered as a small effect, a \(d\) of 0.5 as a moderate effect, and a \(d\) of 0.8 as a large effect\(^{27}\).
Figure 1 presents the flow diagram of participants in the project. Of the 60 eligible first-year GPTs, 53 were allocated to either the intervention group (n=23) or the control group (n=30). Seven GPTs were excluded for the following reasons: a different educational trajectory (n=2); pregnancy leave (n=1); dropping out of the GP vocational training (n=2); and participation in a previous patient feedback programme (n=2).

In the intervention group one GPT was unavailable for the pre-assessment, but did participate in the patient feedback training and follow-up assessment, and one GPT was unavailable for follow-up assessment. There was one protocol violation: one GPT was excused from participating in the programme because of the extra workload. In the control group two GPTs were lost to follow-up: one had dropped out of the GP vocational training, and one was unavailable for the follow-up assessment.

The analysis was carried out according to ‘intention-to-treat’, which means analysis was performed on the data for the originally allocated groups, irrespective of the individual level of participation in the programme. The baseline characteristics of the two groups are presented in Table 1. No adjustment for variables by means of analysis of covariance was needed\(^{28}\).

All baseline assessments were gathered during a three-week period in September 2006. Follow-up assessments were gathered in a three-week period during December 2006 to January 2007 (after three months). These periods corresponded with the sixth and ninth month of the first year of the GP vocational training. The assessment of the consultations was completed in November 2007.
Figure 1. Flow diagram of the controlled trial
GPTs = General Practitioner Trainees; PF = Patient Feedback; GP = General Practitioner

Eligibility

All first-year GPTs (n=60)

Reasons for not participating:
- different educational trajectory (n=2)
- pregnancy leave (n=1)
- earlier participation in PF project (n=2)
- dropped out of GP vocational training (n=2)

Eligibility

Baseline assessments (n=53)

Enrollment

Class I (n=11) Class II (n=12)

Intervention (n=23)

Class III (n=7) Class IV (n=10) Class V (n=13)

Regular training (n=30)

Allocation

Reasons for loss to follow-up:
- not available at time of assessment (n=1)
- dropped out of GP vocational training (n=1)
- not available at time of assessment (n=1)

Follow-up and analysis

Assessments (n=22) Assessments (n=28)
### Table 1. Baseline characteristics of the general practitioner trainees

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Patient feedback (intervention group) (n=23)</th>
<th>Regular training (control group) (n=30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean, SD), years</td>
<td>31.1 (4.1)</td>
<td>31.9 (6.2)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>6 (26.1%)</td>
<td>12 (40.0%)</td>
</tr>
<tr>
<td>female</td>
<td>17 (73.9%)</td>
<td>18 (60.0%)</td>
</tr>
<tr>
<td>Clinical work experience (mean, SD), years</td>
<td>2.8 (2.2)</td>
<td>2.4 (1.6)</td>
</tr>
<tr>
<td>Dutch as first language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>20 (87.0%)</td>
<td>26 (87.7%)</td>
</tr>
<tr>
<td>no</td>
<td>3 (13.0%)</td>
<td>4 (12.3%)</td>
</tr>
<tr>
<td>Practice location area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>large city</td>
<td>8 (34.8%)</td>
<td>8 (26.7%)</td>
</tr>
<tr>
<td>town</td>
<td>11 (47.8%)</td>
<td>12 (40.0%)</td>
</tr>
<tr>
<td>rural</td>
<td>4 (17.4%)</td>
<td>10 (33.3%)</td>
</tr>
<tr>
<td>Clinical knowledge test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pass</td>
<td>20 (87.0%)</td>
<td>24 (80.0%)</td>
</tr>
<tr>
<td>failure</td>
<td>3 (13.0%)</td>
<td>6 (20.0%)</td>
</tr>
</tbody>
</table>

SD = Standard Deviation
### Table 2. Primary outcomes

<table>
<thead>
<tr>
<th></th>
<th>MAAS-Global scores baseline assessments (SD)</th>
<th>MAAS-Global scores follow-up assessments (SD)</th>
<th>change</th>
<th>p</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>All GPTs (n = 53)</td>
<td>3.29 (0.75)</td>
<td>3.54 (0.66)</td>
<td>0.25</td>
<td>p = .047</td>
<td>0.03 – 0.49</td>
</tr>
<tr>
<td>Intervention group (n = 23)</td>
<td>3.25 (0.73)</td>
<td>3.46 (0.57)</td>
<td>0.21</td>
<td>Interaction effect = -0.05, p = .843</td>
<td>-0.54 – 0.44</td>
</tr>
<tr>
<td>Control group (n = 30)</td>
<td>3.33 (0.75)</td>
<td>3.59 (0.73)</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SD = standard deviation; C.I. = confidence interval; GPTs = General Practice Trainees

### Table 3. Participation rate and consultation skills (intervention group)

<table>
<thead>
<tr>
<th>Process measures</th>
<th>n</th>
<th>MAAS-Global scores baseline assessments (SD)</th>
<th>MAAS-Global scores follow-up assessments (SD)</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of PFCs received</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥10</td>
<td>12</td>
<td>3.17 (0.73)</td>
<td>3.41 (0.65)</td>
<td>1</td>
</tr>
<tr>
<td>&lt;10</td>
<td>10</td>
<td>3.43 (0.76)</td>
<td>3.50 (0.54)</td>
<td>0</td>
</tr>
<tr>
<td>Formulated learning points</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥10</td>
<td>13</td>
<td>3.13 (0.63)</td>
<td>3.48 (0.79)</td>
<td>1</td>
</tr>
<tr>
<td>&lt;10</td>
<td>9</td>
<td>3.37 (0.80)</td>
<td>3.43 (0.48)</td>
<td>0</td>
</tr>
<tr>
<td>Discussed topics with GP-trainer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>11</td>
<td>3.08 (0.74)</td>
<td>3.45 (0.68)</td>
<td>1</td>
</tr>
<tr>
<td>No</td>
<td>8</td>
<td>3.66 (0.59)</td>
<td>3.33 (0.44)</td>
<td>0</td>
</tr>
</tbody>
</table>

SD = Standard deviation; PFC = Patient Feedback questionnaire of Consultation skills; GP = General practitioner; Z-score = dichotomised score
Primary outcome measurement
As Table 2 shows, MAAS-Global mean scores for all participants were 3.29 (SD 0.75) at baseline and 3.54 (SD 0.66) at follow-up (p=0.047). Thus the GPTs improved their consultation skills within the study time interval. In the intervention group scores improved from 3.25 (SD 0.73) to 3.46 (SD 0.57), showing very little difference (p=0.84) from scores in the control group, which improved from 3.33 (SD 0.75) to 3.59 (SD 0.73).

Process measures
The mean number of PFCs received per GPT was 10.4 (SD 6.0, range 0-21); 19 (82.6%) GPTs managed to formulate learning points. The mean number of formulated learning points (on the PFC self-assessment version) was 7.8 (SD 6.2, range 0–21). Furthermore, 11 (58%) of the participating GPTs discussed their learning topics with their GP trainer. Table 3 shows the baseline and follow-up assessments for the dichotomised determinants of participation. Trainees with Z-scores of 3 and 2 (active participation) were found to have a small-to-moderate improvement in MAAS-Global scores of 0.32 (SD 0.32), whereas GPTs with Z-scores of 1 and 0 had an improvement of 0.02 (SD 0.25) (p=0.11).

Other determinants of change
Multivariate linear regression analysis identified no determinants (age, gender, clinical work experience, Dutch as first language, practice location area and clinical knowledge skills test) that contributed significantly to the improvement in MAAS-Global scores as a dependent value (the data for the two groups were combined and analysed together).
Discussion

Main results
Consultation skills in the entire cohort of participants improved with a small-to-moderate effect size within the three-month observation period (research question 1). Consultation skills in the group which undertook the patient feedback training programme did not improve any more than those in the control group, which received regular doctor-patient communication education (research question 2). However, a subgroup of GPTs who participated actively in the programme showed a greater improvement in their consultation skills than those who did not actively participate (research question 3).

Interpretation of the results
Although attending the patient feedback programme did not seem to be advantageous in terms of developing consultation skills, we think that it is important to take certain factors into consideration.

Firstly, our GPTs had already achieved a high standard of consultation skills. Our cohort scored of mean of 3.29 on the MAAS-Global instrument at baseline, which is high compared with, for example, the mean of 2.2 achieved by a similar cohort of first-year GPTs on the MAAS-Global instrument. High baseline scores are difficult to improve, which may imply that the effect of the essential element, patient feedback, may be lost in the overall process.

Secondly, we chose a study design with high generalisability (external validity). This means that we studied the patient feedback programme in its natural setting. The research team did not interfere in the actual teaching and monitoring of students, because this might have implications for the external validity of the results. The sub-group results indicate that stricter implementation would perhaps have given better results. Only a minority of the GPTs managed to acquire patient feedback from the 20 patients recommended for generalisability. Dichotomising GPTs in subgroups according to whether they were ‘actively participating’ or ‘non-actively participating’ showed improvement in the consultation skills of the former group, but we cannot
conclude that patient feedback is the only determinant accountable for this effect because other undefined covariates may be responsible.

Thirdly, the educational impact of patient feedback might be limited by the relatively positive responses that are often given by patients\textsuperscript{10,13}, possibly as a result of the unequal balance of power between patients and doctors. We developed our patient feedback programme and the PFC in such a way as to optimise the patient feedback potential by, for example, stimulating GPTs to select and recruit patients from whom they expected to obtain relevant feedback\textsuperscript{19}. Other methods, however, such as more detailed verbal feedback from a sub-sample of patients (such as might be acquired by obtaining patients’ synchronised comments to audio or videotaped consultations), tailored coaching sessions with a teacher in which the collated results of surveys are discussed, or combined methods, may make the educational impact of patient feedback interventions stronger and help to triangulate and add context to the survey responses.

**Context of existing literature**

In view of the dearth of studies that assess the effect of patient feedback on consultation skills, our results are relevant, but are also difficult to correlate. The fact that consultation skills improved during a period of structured education is in line with the findings of a previous study, in which junior doctors in particular showed improvement\textsuperscript{29}, but somewhat contrasts with the results of long-term structural vocational education in which the development of communication skills did not reach a level of significance\textsuperscript{3}. The limited effect of our patient feedback intervention is in line with the results of a study of GPs in the Netherlands\textsuperscript{15}, but contrasts with the results of a study in which a beneficial effect on consultation skills of GPTs was demonstrated\textsuperscript{23}.

**Strength and limitations**

Our study is strengthened by the involvement of SPs. The random procedure for assessing the videotaped consultations probably also has contributed to the reliability of the results. A possible alternative would have involved analysing the direct results of the patient feedback questionnaires, which carries the risk of a
selection bias. In our protocol, GPTs were able to select those patients from whom they expected to receive relevant feedback, which, in our opinion, has educational advantages. Furthermore, the fact that our sample included two homogeneous sub-cohorts of first-year GPTs, which were subject to close monitoring and which suffered limited loss to follow-up, will also have contributed to the representativeness of the results.

A limitation of our study is that the setting of class education did not allow randomisation at participant level. Cluster randomisation with only five groups would imply the risk of a misbalanced allocation and was therefore not considered. We were of the opinion that the (blinded) allocation of two sub-cohorts would thus be optimal.

Another methodological limitation of this study concerns the limited size of the cohort. Although the difference between baseline and follow-up assessment results exceeded the level of significance, the effect size was small to moderate. We conclude that the power in this study was relatively low. However, we chose for the homogeneity of a cohort of first-year GPTs from one training institute.

The scoring of multiple different consultations has been recommended for individual assessment of consultation skills\textsuperscript{25}. For logistical and financial reasons we chose to assess one baseline consultation and one follow-up consultation with SPs by two independent and blinded assessors.

**Conclusion**

The concept of using patient feedback for learning purposes remains intriguing. That we were unable to demonstrate an increase in beneficial effects on consultation skills by our patient feedback programme over regular communication skills training in this study by no means implies that patient feedback is less relevant. To study learning processes in relation to input from trainers, patients and researchers has proven to be a complex endeavour. Feedback from the target group, the patients, might have other implications.
beyond the field of consultation skills, such as aspects of satisfaction with the consultation or practice organisation. We consider the provision of diverse and integrated feedback to GPTs to be essential to their development of a professional attitude and the necessary skills, and emphasize the importance of exploring the possibilities of education in this specific field.
References


21. Cheraghi-Sohi S, Bower P. Can the feedback of patient assessments, brief training, or their combination, improve the interpersonal skills of primary care physicians? A systematic review. BMC Health Serv Res 2008;8:179.


The educational potential of patient feedback on physicians’ consultation skills: a systematic review

Marcel E. Reinders
Bridget L. Ryan
Annette H. Blankenstein
Henriëtte E. van der Horst
Moira A. Stewart
Harm W.J. van Marwijk

Submitted
Abstract

Objective
The effectiveness of patient feedback assessments as a method of improving physicians’ consultation skills is equivocal. Research is scarce and shows a wide variety in methods and rigour. In this systematic review, we analysed the evidence for the educational potential of real-patient feedback on consultation skills of practicing physicians.

Methods
Five electronic databases were searched (PubMed, EMBASE, Cochrane, PsycInfo and Eric) and screened against eligibility criteria: 1) practicing physicians, 2) real patient feedback, 3) general consultation skills, 4) general health care, 5) physicians being actually confronted with individual feedback. Empirical studies of all study designs (randomized (controlled), quasi-experimental, cross-sectional and qualitative designs) were eligible for inclusion. Two authors selected the articles for inclusion independently against eligibility criteria. Articles were assessed for level of educational impact: views on learning experience (level 1); a change in knowledge or skills (level 2); an intended change in the physicians’ performance (level 3); and an actual change in performance (level 4), and quality according to a best evidence medical education (BEME) coding scheme.

Results
Fifteen studies were identified (10 studies in primary care and five in other specialties) conducted in the United States, the Netherlands, the United Kingdom, Australia and Canada. The method of tailored reports with aggregated patient feedback results for transferring feedback was most commonly described. Eighty percent of the studies showed an improved outcome (effect) as a result of patient feedback. However, positive results were more common among studies assessing, for example, perceptions of participants’ knowledge (considered a low level of educational impact) than studies assessing actual
change in consultation skills (considered the highest level of educational impact) where four out of seven studies found beneficial effects.

**Conclusions**

There is evidence for the educational potential of patient feedback on physicians’ consultation skills, however, the evidence for actual improvement is limited. The implication therefore is that the rationale for implementation of patient feedback into practice depends on the level of aimed impact.
Introduction

The importance of patient assessments and feedback as a stimulus for improving physicians’ consultation skills, or as an incentive for efforts to improve quality of medical care, is gaining increased attention\(^1,2\). Considering patients’ views on the delivery of medical care seems to be self-evident and is also related to improved patient health outcomes and satisfaction\(^3-5\). The therapeutic essence of a doctor-patient relationship should include the patient perspective obtained either from ratings or surveys after encounters\(^1\). However, to what extent patient feedback is effective in improving physician performance remains equivocal.

Patient feedback on communication or interpersonal skills of physicians commonly involves assessments by means of surveys\(^6,7\). This feedback is then transferred to physicians through a variety of methods such as tailored reports or coaching sessions, in which the collated results of patient surveys are presented, or through focus groups in which patient feedback assessments are discussed. Patients appear comfortable with the various methods of involvement\(^8\). Formal feedback on the assessment of physicians from various sources (e.g. mentors, trainers, colleagues) is effective in improving clinical performance\(^2\), but studies lack rigorous methodology\(^6\). The evidence for the effectiveness of feedback assessments from patients is scarce and seems to be hampered by the heterogeneity of studies\(^9,10\) and the scarcity of rigorously validated instruments\(^7,11\). In two recent systematic reviews about instruments and patient feedback methods\(^12\), and about patient feedback assessments and brief training interventions in general practice\(^13\), no conclusive evidence was found for the effectiveness of patient feedback interventions. Additionally, these reviews only included randomized trials, which are scarce in this field. What is still missing, therefore, is a review of studies that focuses on other aspects of educational potential of patient feedback interventions, in addition to those that focus on actual change in skills. Due to the complex nature of the effectiveness and educational potential of patient feedback, these studies are often conducted in various, non-randomized designs\(^6,14\). The effectiveness of patient feedback on
consultation skills is found on different levels of educational impact. In the Kirkpatrick hierarchy\textsuperscript{15,16}, four levels on which educational interventions can have impact are identified: 1. views on learning experience; 2. change in knowledge or skills; 3. intended change in performance; and 4. actual change in performance.

In this systematic review, we addressed the following research question: What is the evidence for the educational potential (from physicians’ valuation of learning experience to physicians’ behavioural change) of individually directed, real-patient feedback on general consultation skills of practicing physicians? Therefore, we included all studies (randomized trials, or quasi-experimental studies and qualitative research) with different levels of aimed educational impact. We explored participant appreciation and physician behavioural change, assessed by patients and/or by physicians. Furthermore we performed a quality assessment of included studies and made suggestions for practical implications and for future research.

**Methods**

**Search strategy**

The search strategy, based on the search definition ‘physicians receiving feedback of real patients on their consultation skills’, used the following controlled vocabulary and free text words: ‘feedback’, ‘clinical competence’ and ‘health care’ (full search strategy available from MR). A medical information specialist applied the search (last update on April 23\textsuperscript{rd} 2009), to the following databases: PubMed, EMBASE, Cochrane, PsycInfo and Eric, without restrictions to language or publication year. Other sources for eligible articles were reference lists of related systematic reviews identified in this search, and the articles included for final analysis.
Eligibility criteria

We included all empirical studies regardless of study designs meeting the following eligibility criteria:

- practicing physicians, including physicians in postgraduate training (but not medical students)
- real patient feedback (but not standardized simulated patients or other parties, like health workers, nurses or peers only, giving feedback)
- general consultation skills, (but not satisfaction of received care or practice organization, or articles limited to one aspect of the consultation, e.g. giving bad news, shared decision-making, palliative care)
- feedback on communicative aspects in general health care (but not only related to a specific disease, or to patients or physicians concerned with a specific disease)
- physicians being actually confronted with individual feedback (but not conceptual studies of the value of patient feedback or validation studies of surveys, without feedback to physicians).

Selection process

One reviewer (MR) screened the titles and abstracts of all citations to determine those matching the search definition (first step, ‘low-threshold’ strategy). Two independent authors (MR, BLR) further selected articles on the basis of title and abstract fulfilling the eligibility criteria (second step, ‘eligibility criteria’), which led to a selection of articles for full text review. Disagreement about articles fulfilling eligibility criteria was resolved by discussion. Full text articles that met the eligibility criteria were included for final analysis (third step, ‘final selection’).

Data extraction and analysis

For all included studies, two reviewers (MR, BLR) performed data extraction (two of the studies were written by the first author; these articles were reviewed by an independent assessor). Discrepancies were resolved by discussion. The template used in this study is based on the standard best evidence medical education (BEME) organisation coding sheet\textsuperscript{17}, and contains the following areas:
(1) aim of the study, (2) study design, (3) study population, (4) intervention type and intensity of patient feedback interventions, (5) level of educational impact according to the Kirkpatrick hierarchy: views on learning experience (level 1); a change in knowledge or skills (level 2); an intended change in performance (level 3); and an actual change in the physicians’ performance (level 4); (6) learning outcomes; (7) effectiveness of patient feedback; and (8) quality assessments. We chose not to pool results of the studies, because we expected beforehand that studies would be too heterogeneous methodologically. In Figure 1 a schematic overview is presented of the several steps from patient feedback assessments to outcome assessments of the learning effects.

**Quality assessment**

Quality assessments were conducted separately for quantitative research and qualitative research. Studies with mixed methods were analysed according to the core method that was used. Quality assessments of quantitative studies were based on four major items: prevention of bias (participant selection, patient selection, adherence to the intervention and validity of the instruments), participant sample size, intensity of the intervention (number of involved patients, tailored coaching sessions and duration of the intervention), and the appropriateness of the intervention and adequacy of description (based on a standard sheet to extract the quality of clinical trials)\(^{18}\). Quality assessments of qualitative studies were based on: well defined goals, appropriateness of the design, the appropriateness of data collection, and rigour of conduct, analysis and reporting\(^{19}\). Studies were not excluded for analysis because of low quality scores.

**Classification of study effects by study characteristics**

The outcome of the effects on consultation skills (classified as positive or negative) as a result of patient feedback was related to: intervention type, aimed level of impact, assessment level (patients or physicians), stage of education, setting (country, medical specialty), and publication year.
Figure 1. A model of patient assessments, feedback procedure and outcome assessments

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Method</th>
<th>Subject of Study</th>
<th>Assessment of Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>real-patient assessments (surveys, interviews)</td>
<td>questionnaires</td>
<td>intentions of behavioural change</td>
<td>self-assessment</td>
</tr>
<tr>
<td></td>
<td>tailored reports</td>
<td>modification of knowledge or skills</td>
<td>patient assessment</td>
</tr>
<tr>
<td></td>
<td>small group discussions</td>
<td>valuation or appreciation</td>
<td>teacher/researcher assessment</td>
</tr>
</tbody>
</table>
Results

Search
Figure 2 shows a flow-diagram of the results of the selection process. After removing 815 duplicates, a first broad screening of 8452 citations (8410 from electronic data-bases and 42 from hand-search) on titles and abstracts matching the search definition yielded 287 possible articles. After a second screening on title and abstracts against eligibility criteria, 108 full text articles were retrieved, of which 15 studies met the eligibility criteria and were included\(^\text{20-34}\). Two studies were based on the same data-set\(^\text{33,34}\) but explored different outcome measures and were therefore both included.

Reasons for exclusion were: 42 patient assessments studies but no actual feedback\(^\text{8,35-75}\); 26 validation studies of patient assessment questionnaires\(^\text{76-101}\); five non-empirical studies\(^\text{102-106}\); 12 reviews \(^\text{1,3,9,10,12,13,107-112}\); and eight studies involved no real patients, no reflection on individually addressed general consultation skills, or involved undergraduate students\(^\text{113-120}\).

Description of Studies
Table 1 describes each of the included studies.

Setting
The included studies were conducted in the USA\(^\text{21-24,32}\), the Netherlands\(^\text{30,31,33,34}\), United Kingdom\(^\text{20,27,29}\), Australia\(^\text{26,28}\), and Canada\(^\text{25}\). Ten studies concerned post-graduate training (general practice\(^\text{24,26-28,30,31}\), internal medicine\(^\text{22,23,32}\), and paediatrics\(^\text{21}\)); five studies concerned professional education (general practice\(^\text{25,29,33,34}\) and orthopaedics\(^\text{20}\)).

Instruments for acquiring patient feedback
All but one\(^\text{29}\) study made use of patient feedback questionnaires (mostly validated), focusing on individual physician performance. Only the Patient Satisfaction Questionnaire was used by more than one study group\(^\text{21,23,32}\). Two
studies made use of structured patient interviews\textsuperscript{24,29}, one study did this in combination with written questionnaires\textsuperscript{24}.

**Patient feedback presentations**

The commonest method of transferring feedback from patients to physicians was by means of aggregated/collated results of questionnaires in tailored reports made by preceptors/researchers\textsuperscript{20,24-28,32-34}. Another method was tailored coaching, in which the feedback results were mediated by a researcher/preceptor, varying from one single coaching session to quarterly meetings\textsuperscript{21-24}. The method of consultation-specific feedback questionnaires directly presented to the physician was applied in two studies\textsuperscript{30,31}. In two studies, small group sessions discussing patient feedback were conducted\textsuperscript{27,29}, either organized monthly during one year, or once at the end of the study period.

**Additional measures to enhance the educational potential of patient feedback**

Some studies used physician self-assessment reports of their consultation skills, in addition to patient feedback questionnaires. This challenged the participants to reflect on their performance and the learning possibilities. Physicians’ self-assessment reports were made after each single consultation\textsuperscript{30,31}, or as a concluding report\textsuperscript{20,21,26-28}. The possibility of discussing the results and meaning of patient feedback with a preceptor\textsuperscript{27} or GP trainer\textsuperscript{29-31} was meant to further intensify the learning effect.

**Outcomes of Studies**

**Outcome measures and their assessments**

The outcome measures were defined as: consultation skills\textsuperscript{30,31}, communication skills/behaviour\textsuperscript{21,23-25,34}, interpersonal skills\textsuperscript{20,26-28}, humanistic qualities\textsuperscript{32}, and non-technological skills\textsuperscript{22}. Communication skills and interpersonal skills form an integrated competence\textsuperscript{1}. Outcome measures were assessed by patients\textsuperscript{23,33}; by physicians\textsuperscript{25,32}, by a third party\textsuperscript{24,29,30,34}, or in combination\textsuperscript{20-22,26-28,31} (Figure 1).
Study designs and the intended level of educational impact

Figure 3 shows a model in which the level of educational impact is integrated with the study design (qualitative or quantitative research). The four randomized controlled trials all aimed at level 4 impact (actual change). The quasi-experimental studies aimed at various levels of impact (from level 4 to level 1, of which one study (Fiddler) explicitly aimed at both level 3 (intended change) and level 4). Of the four cross-sectional studies and the two qualitative studies, the aimed impact was at level 2 or level 1.

Effect size (data synthesis)

Twelve of the 15 studies (80%) showed improved outcome (effect) as a result of patient feedback. Table 2 describes the outcomes and significance for each study. The effects by specified level of educational impact were:

- Level 4 (observed change):
  Brinkman et al described 35% increase of ratings of consultation skills as a result of patient feedback\textsuperscript{21} (outcome reported by nurses and not by patients’ (parents) assessments). Fiddler et al described that approximately 12% of physicians initiated changes on communication and support of patients\textsuperscript{25} (self-assessments). Cope et al described a ‘large effect’ on improved patient valuation of consultation skills\textsuperscript{23}, but Greco et al found only a small effect at the upper end of a patient satisfaction scale\textsuperscript{26}. In contrast, in the studies of Vingerhoets and Wensing, and of Reinders, no difference in improved consultation skills was reported\textsuperscript{30,33,34} (patients and preceptor assessments), however, in the latter study sub-group analysis showed that active adherence to the study programme was related to more improvement in consultation skills\textsuperscript{30}.

- Level 3 (intended change):
  Fiddler et al\textsuperscript{25} reported that 19% of physicians contemplated a change in communication and patient support as a result of patient feedback (versus 12% initiated change, see also level 4).
• Level 2 (modification of knowledge or skills):
In four studies (Al-shawi, Greco (1995, 2001) and Radcliffe), a majority of participants self-reported a modification in knowledge or skills as a result of patient feedback\textsuperscript{20,27-29}, especially if the learning instrument gave insight into their deficiencies\textsuperscript{27}. Among issues that were identified as important for change were listening, explaining and investing time\textsuperscript{29}, awareness and more concern with the patients\textsuperscript{22}, or communication skills in general\textsuperscript{20,28}.

• Level 1 (appreciation/valuation):
Falvo et al described that patient feedback was considered by physicians as a useful reflection on strengths and weaknesses\textsuperscript{24}. In the study of Thomas\textsuperscript{32}, only patient feedback on humanistic qualities was considered relevant. In the study of Reinders\textsuperscript{30}, the usefulness of patient feedback was limited to the more complex consultations (content related). The high patient scores suggested to the physicians that patients were providing socially desirable answers. Physicians viewed this as unhelpful and this hampered physicians’ adherence to the programme. Some quantitative studies also presented valuation data but this was not the main focus of these studies. For example, Greco et al described that GP registrars perceived better understanding of interpersonal skills\textsuperscript{26}, whereas in the study of Wensing\textsuperscript{34}, physicians found patient feedback to be time-consuming and difficult to learn from (due to high-levels of patient satisfaction).

Quality of the included studies
Table 3 shows the quality assessments of the quantitative studies (Table 3 a) and the qualitative studies (Table 3 b), separately.

Classification of study effects by study characteristics
Positive changes were found throughout the professional stages (experienced clinicians as well as postgraduates). Positive change did not depend on the intervention method (structured reports, tailored coaching, or otherwise), or assessment level (patients, participants). All studies that aimed at lower level of educational impact (1 and 2) found positive results; at impact level 4, however,
Context of existing literature (systematic review)

only four out of seven studies found a positive result. Positive outcomes were demonstrated in studies conducted in English-speaking countries (throughout the last three decades, among all specialties), whereas interventions were shown ineffective, as described in the Netherlands (in the last decade, in general practice).

Figure 2. Flow-diagram of study selection

Identification
- 9225 records through database screening
- 42 additional records identified through hand search

Screening
- 8452 records after duplicates removed screened on title and abstract
- 8344 records excluded

Eligibility
- 108 full-text articles assessed for eligibility
- 93 full-text articles excluded

Included
- 15 studies included
<table>
<thead>
<tr>
<th>Study</th>
<th>Study aim</th>
<th>Intervention</th>
<th>Study design</th>
<th>Setting and population</th>
<th>Level/stage of education</th>
<th>Patients per physician</th>
<th>Patients per group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brinkman (2007)</td>
<td>Assessment of effectiveness of multisource feedback on consultation skills and professionalism</td>
<td>Tailored coaching with collated results of parent, nurse and self-assessment evaluations</td>
<td>Randomized (controlled) trial</td>
<td>18 pediatric residents (18 controls), USA</td>
<td>Postgraduate training (PT)</td>
<td>18</td>
<td>Not specified</td>
</tr>
<tr>
<td>Brody (1980)</td>
<td>Assessment of effectiveness of a project to improve recognition of non-technological aspects of medical care</td>
<td>Tailored coaching with collated results of patient interviews</td>
<td>Qualitative, experimental</td>
<td>43 residents internal medicine (9 controls), USA</td>
<td>PT</td>
<td>4-6</td>
<td>≥ 6</td>
</tr>
<tr>
<td>Cope (1986)</td>
<td>Assessment of effectiveness of a project to improve recognition of non-technological aspects of medical care</td>
<td>Tailored coaching with collated results of patient interviews</td>
<td>Quasi-experimental, with control group</td>
<td>9 residents internal medicine (9 controls), USA</td>
<td>PT</td>
<td>≥ 6</td>
<td>4</td>
</tr>
<tr>
<td>Falvo (1980)</td>
<td>Assessment of feasibility and educational value of patient feedback as a tool for developing patient teaching skills</td>
<td>Tailored coaching with collated results of patient surveys</td>
<td>Qualitative analysis (interviews)</td>
<td>7 GP residents, USA</td>
<td>PT</td>
<td>≥ 6</td>
<td>50</td>
</tr>
<tr>
<td>Fidler (1999)</td>
<td>Assessment of initiated changes as a result of patient evaluation feedback as a tool for developing patient teaching skills</td>
<td>Tailored coaching with collated results of patient surveys</td>
<td>Quasi-experimental, with control group</td>
<td>218 GP residents, Canada</td>
<td>PT</td>
<td>25</td>
<td>2850</td>
</tr>
<tr>
<td>Greco (2001), Greco (1995), Al-Shawi (2005)</td>
<td>Feasibility of introducing patient feedback into the surgical training of GPs or into the professional training of surgeons</td>
<td>Collated summary of patient surveys; additional preceptor feedback discussion</td>
<td>Quasi-experimental, 2 control groups</td>
<td>33 GP residents, Australia, 13 pairs of GP residents and GP trainers, UK, 10 orthopedic surgeons, UK</td>
<td>PT</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>Al-Shawi (2005)</td>
<td>Assessment of impact of different models of patient feedback on interpersonal skills development</td>
<td>Individualised structured feedback report of collated results of patient surveys</td>
<td>Cross-sectional, both quantitative and qualitative</td>
<td>28 pairs of GP trainers, UK</td>
<td>PT</td>
<td>25</td>
<td>10</td>
</tr>
</tbody>
</table>

**Table 1. Description of included studies**
<table>
<thead>
<tr>
<th>Study</th>
<th>Study aim</th>
<th>Intervention</th>
<th>Study design</th>
<th>Setting and population</th>
<th>Patients per physician</th>
<th>Level/stage of education</th>
<th>Validated patient feedback instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radcliffe (1997)</td>
<td>Reflection on invoke of patients’ accounts of care as audit tool</td>
<td>Small group discussions about the results of patient interviews</td>
<td>Qualitative analysis (group interviews)</td>
<td>±40 GPs, UK</td>
<td>10</td>
<td>PE</td>
<td>Semi-structured interviews</td>
</tr>
<tr>
<td>Reinders (2008)</td>
<td>Development of a patient feedback programme as a tool for improvement of consultation skills</td>
<td>Experiential learning with consultation-specific patient feedback questionnaires</td>
<td>Cross-sectional, both quantitative and qualitative</td>
<td>58 GP residents, The Netherlands</td>
<td>20</td>
<td>PT</td>
<td>Patient Feedback Questionnaire on Consultation Skills (PFC)</td>
</tr>
<tr>
<td>Reinders (2009)</td>
<td>Assessment of effectiveness of patient feedback on consultation skills of GP trainees</td>
<td>Experiential learning with consultation-specific patient feedback questionnaires</td>
<td>Cluster-randomized controlled trial</td>
<td>23 GP residents (30 controls), The Netherlands</td>
<td>18</td>
<td>PT</td>
<td>PFC</td>
</tr>
<tr>
<td>Thomas (1999)</td>
<td>Assessment of valuation of patient feedback</td>
<td>Collated summary of patient surveys</td>
<td>Quasi-experimental, with control group</td>
<td>77 residents internal medicine (including controls), USA</td>
<td>7</td>
<td>PT</td>
<td>PSQ</td>
</tr>
<tr>
<td>Vingerhoets (2001)</td>
<td>Assessment of effectiveness of patient feedback on improvement of care</td>
<td>Individualised structured feedback report of collated results of patient surveys</td>
<td>Cluster-randomized controlled trial</td>
<td>30 GPs (30 controls), The Netherlands</td>
<td>30</td>
<td>PE</td>
<td>Chronicle Ill Patients Evaluate General Practice (CEP)</td>
</tr>
<tr>
<td>Wensing (2003)</td>
<td>Assessment of effectiveness of patient feedback on communication behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Table 2. Outcomes

<table>
<thead>
<tr>
<th>Study</th>
<th>Outcome measures</th>
<th>Assessments</th>
<th>Significance (documented improvement or valuation)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Impact level 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brinkman (2007)</td>
<td>Physicians’ communication skills</td>
<td>Parent feedback questionnaires, nurse evaluation questionnaires</td>
<td>Nurse evaluations, but not parent evaluations, showed 35% change in communication with family</td>
</tr>
<tr>
<td>Cope (1986)</td>
<td>Patient satisfaction (friendliness and respect; understanding and non-judgmental attitude; communication skills; technical quality of care)</td>
<td>Parent feedback questionnaires</td>
<td>A ‘large’ effect on improved patient satisfaction (art of care and technical quality)</td>
</tr>
<tr>
<td>Greco (2001)</td>
<td>Interpersonal skills: listening, reassurance, eliciting concerns and fears, time given</td>
<td>Patient feedback questionnaires; Follow-up survey on residents</td>
<td>A small improvement was found for: listening skills, time given to patients, and eliciting concerns and reassurance. Intensity of the patient feedback module was related to improvement in interpersonal skills. 63.4% of the residents had intentions to change behaviour</td>
</tr>
<tr>
<td>Reinders (2009)</td>
<td>MAAS-global score of general consultation skills</td>
<td>Preceptor observation of video-taped consultations</td>
<td>No change in improved consultation skills, although sub-group analysis showed moderate improvement in actively participating GP residents</td>
</tr>
<tr>
<td>Vingerhoets (2001)</td>
<td>General dimensions of care (patient relation and communication; information and advice; support)</td>
<td>Patient evaluation questionnaires</td>
<td>No change in the dimensions of care GPs self-reported changes in their professional performance and organisation of care (no empirical data shown)</td>
</tr>
<tr>
<td>Wensing (2003)</td>
<td>MAAS-global score of general consultation skills</td>
<td>Researcher observation of video-taped consultations; Post-intervention survey on GPs</td>
<td>No change in consultation skills GPs are open minded to learn from patients views, but find it difficult to meet all patients’ needs</td>
</tr>
<tr>
<td><strong>Impact level 3 (and 4)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fidler (1999)</td>
<td>Contemplated or initiated changes in relation to patient evaluations (communication and support of patients; office staff and systems)</td>
<td>Follow-up survey on participating GPs</td>
<td>Feedback by patients, rather than by peers, was most likely to induce initiated changes; especially among physicians with the lowest patient ratings. 12.3% of physicians initiated change on communication and support of patients (19.3% contemplated a change)</td>
</tr>
</tbody>
</table>
Table 2. Outcomes (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Outcome measures</th>
<th>Assessments</th>
<th>Significance (documented improvement or valuation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-Shawi (2005), Greco (1995), Greco (2001)</td>
<td>Physicians’ report on strengths and weaknesses of interpersonal skills and intended actions</td>
<td>Patient feedback questionnaires; Physician self-reports (after receiving feedback); Group discussions with participants</td>
<td>60% of the orthopaedic surgeons, 90% of the GP residents, and 67% of the GPs (highly) valued patient feedback. Some of the surgeons, 75% of the residents and 67% of the GPs tried to modify their interpersonal skills</td>
</tr>
<tr>
<td>Brody (1980)</td>
<td>Patient satisfaction and attitude towards illness</td>
<td>Patient interviews (including Patient Satisfaction and General Health questionnaires); Assessing physicians’ reports of consultations and chart reviews; Physicians evaluation questionnaire</td>
<td>76% of the physicians reported change in communication style and content; 89% received new and unknown information. No change in patient satisfaction or in recognition of non-technological problems. 79% of the physicians reported possibilities for improved patient care, 85% valued the educational experience</td>
</tr>
<tr>
<td>Falvo (1980)</td>
<td>Communication and interpersonal skills (needs were met; listening skills and understanding)</td>
<td>Authors’ personal interviews with participating residents</td>
<td>Residents found the project to be helpful in identifying their strengths and weaknesses in communicating skills</td>
</tr>
<tr>
<td>Radcliffe (1997)</td>
<td>Patients evaluations and GPs reactions on: communication and explanation, doctor-patient relationship</td>
<td>Transcripts of group meetings with GPs</td>
<td>Individual examples of valuations of patient feedback by the participants</td>
</tr>
<tr>
<td>Reinders (2008)</td>
<td>Patients’ perception of patient-centered consultation skills</td>
<td>Patient feedback questionnaires; Group discussions and evaluation forms of GP trainees</td>
<td>45.6% of the residents found patient feedback to be beneficial for improving consultation skills</td>
</tr>
<tr>
<td>Thomas (1999)</td>
<td>Evaluation of patient feedback by residents, concerning humanistic qualities and clinical skills</td>
<td>Follow-up survey on residents</td>
<td>A sustained valuation of patient feedback on humanistic qualities; but a decline in valuation of patient feedback on other clinical skills</td>
</tr>
</tbody>
</table>
Figure 3. An integrated model of research type (quantitative, qualitative) and level of educational impact (Kirkpatrick’s hierarchy level 1-4)
### Table 3 a. Quality of quantitative studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Prevention of bias</th>
<th>Sample size participants</th>
<th>Intensity of intervention</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>Al-Shawi (2005)</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>Brinkman (2007)</td>
<td>1.0</td>
<td>0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Brody (1980)</td>
<td>0</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Cope (1986)</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Fidler (1999)</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>Greco (1995)</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0</td>
</tr>
<tr>
<td>Greco (2001)</td>
<td>0.5</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Reinders (2009)</td>
<td>1.0</td>
<td>0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Thomas (1999)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>Vingerhoets</td>
<td>1.0</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>(2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wensing (2003)</td>
<td>1.0</td>
<td>0</td>
<td>0.5</td>
<td>1.0</td>
</tr>
</tbody>
</table>

A: randomized group of participants; B: randomized group of patients; C: controlled for adherence of participants to the intervention; D: use of validated instruments; E: sample size participants > 25; F: number of involved patients per physician > 10; G: tailored coaching of patient feedback intervention; H: duration of the intervention > 6 months; I: appropriateness of intervention; J: well described and equivocal presentation of results and conclusions; 1.0 = yes; 0 = no; 0.5 = description was unclear or missing or only partly met our criteria.

### Table 3 b. Quality of qualitative studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Well defined goals</th>
<th>Appropriateness of design (focus groups)</th>
<th>Appropriateness of data collection</th>
<th>Rigour of conduct</th>
<th>Rigour of analysis</th>
<th>Rigour of reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Falvo (1980)</td>
<td>1.0</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>Greco (2001)</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>Radcliffe (1997)</td>
<td>0.5</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Reinders (2008)</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
</tr>
</tbody>
</table>

1.0 = yes; 0 = no; 0.5 = description was unclear or missing or only partly met our criteria.
We have identified 15 studies in which the educational potential of real patient feedback on general consultation skills has been investigated. These studies cover three decades of research and vary in study design, levels of educational impact, and quality. A majority of the 15 studies observed a positive effect of real patient feedback on consultation skills of physicians; however, there are some caveats. First, the positive results are clustered among studies using an outcome measure at the low end of Kirkpatrick’s hierarchy of educational impact, and were non-randomized and qualitative studies. Where the impact level was high in the hierarchy (actual change in consultation skills), four of seven studies found a positive effect. Second, only one of four randomized trials found positive results.

A possible reason for this phenomenon is that the assessment of actual change in general consultation skills can be difficult, due to the lack of precision in defining consultation skills and the lack of responsiveness of the outcome instruments. Greco et al report that changes at the high end scale (i.e. near the ceiling) of a measuring instrument and a limited effect size are related. Baseline scores at the high end of the scales, a concern in most of the studies, appear to make changes in behaviour less urgent, when in reality these high scores may reflect a difficulty in measurement rather than true high satisfaction with physician communication. An important consideration in measurement is that learning (and talking) is a subjective process. Authenticity and person-centeredness can be objectified and counted, but perhaps in the process of transferring them from the subjective to the objective, some of meaning, validity and relevance are lost.

In this review, some correlations between study characteristics were studied. The assumption that the intensity of patient feedback interventions was related to increased changes in outcomes, could not be confirmed or rejected, mainly because of the plethora of the applied methods and the lack of description of the adherence of participants to the programme. Furthermore, it is not clear why in the English-speaking countries (USA, Canada, United
Kingdom, Australia) more positive results were found, in contrast to the studies in the Netherlands, although of course, a relatively small number of studies is involved. Moreover, there were no indications that younger doctors (residents) were more liable to change than more experienced clinicians, who might have acquired routine or fixed patterns of consultation. This suggests that patient feedback may be as useful in faculty development and continuing medical education as it is in post-graduate education.

In contrast to the vast number of patient satisfaction studies, effect studies of patient assessments are less common; and randomized controlled trials on the educational impact of patient feedback are a rarity. Finding evidence for the educational value of the assessments and feedback assessments of practicing physicians was hampered by the heterogeneity of studies. This heterogeneity made effect sizes difficult to standardize and compare. The possibilities of an integrated presentation of qualitative research have been demonstrated by several others. In this review, we focussed on the aimed level of educational impact of the included studies, as a way of a combined and integrated presentation of the qualitative and quantitative studies.

Future patient feedback effect studies should focus on the highest level of educational impact, that of actually achieved changes in consultation skills. Qualitative studies with high rigour are also needed, in order to achieve: better understanding of how physicians can use patient feedback to improve communication; valid assessment of participants (patients and physicians, and teachers); preferred methods in which patient feedback should be acquired and presented; and high levels of adherence to training programmes.

**Conclusion**

Beyond the principle that patient involvement in health care is self-evident, and morally justifiable, we need a better understanding of how physicians can use patient feedback to learn to communicate better with their patients. This review has demonstrated that there is evidence for the effectiveness of real patient feedback interventions on lower levels of educational impact such as perceived
change in knowledge, but few studies and little evidence for an impact on the higher level such as actual change physicians’ consultation skills.
References


5. Stewart MA. Effective physician-patient communication and health outcomes: a review. CMAJ 1995 May 1;152(9):1423-33.


13. Cheraghi-Sohi S, Bower P. Can the feedback of patient assessments, brief training, or their combination, improve the interpersonal skills of primary care physicians? A systematic review. BMC Health Serv Res 2008;8:179.

Chapter 6


36. Branch WT, Malik TK. Using 'windows of opportunities' in brief interviews to understand patients' concerns. JAMA 1993 April 7;269(13):1667-8.


CHAPTER 7

Who is afraid of patient feedback?
Analysis of self-directed learning of consultation skills among general practice trainees

Marcel E. Reinders
Annette H. Blankenstein
Harry Schleypen
Henriëtte E. van der Horst
Harm W.J. van Marwijk

Submitted
Summary

There is evidence for the beneficial effects of including the opinions of patients in primary health care. Therefore, in the Institute for vocational training in general practice at the VU medical center in the Netherlands, we introduced a patient feedback training programme. General practice trainees (GPTs) could learn from their patients by asking for feedback by means of a validated questionnaire. However, wide variances in participation were found among the GPTs (i.e. the number of feedback questionnaires they collected).

In this mixed-method study, we analysed whether 22 first-year GPTs were apprehensive of greater patient involvement in their education, and studied determinants that predicted participation in the programme.

We conducted semi-structured interviews with six GPTs who collected less than five feedback questionnaires from their patients. These GPTs did not feel vulnerable when they asked patients for feedback, neither did they question the benefits of patient involvement in their education, but had various motivational and practical reasons for their low rate of participation.

It is not clear to what extent GPTs recognize their learning needs, especially in a ‘self-directed’ learning programme with a sensitive topic, such as feedback from real patients. The baseline level of consultation skills did not predict the rate of their participation in the programme. Among male GPTs, and GPTs with more clinical experience, there was a lower rate of participation.

Close monitoring of GPTs by tutors might help the GPTs to overcome practical problems. This study has provided more insight into further requirements in the development of effective self-directed learning programmes with patient involvement.
**Introduction**

Methods to include patients’ views on general practice care are considered to be important for the improvement of care\(^1,2\), and are related to health outcomes as well as patient satisfaction\(^3,4\). Feedback from patients on a doctor’s performance can be useful in developing patient-centred consultation skills\(^5\), which is a core element in the education of primary care professionals\(^6\), and the communication style preferred by patients\(^7\).

The reactions of some general practitioners to patient assessment of their communication skills have been reported to be positive\(^8,9\) but others were less positive\(^10\). Doctors, with or without clinical experience, might feel awkward about asking their patients for feedback. Especially for general practice trainees (GPTs), patient feedback can be daunting, and their first reactions can be defensive, but many GPTs react favourably to knowing what their patients think about them\(^11\).

In 2006, the patient feedback training programme was introduced in our vocational training education for general practice (VU medical center, the Netherlands), to create opportunities for students who want to learn from their patients\(^12\). GPTs handed out feedback questionnaires on their consultation skills to patients, directly after a consultation. What we soon noticed was that the rate of participation of the GPTs varied widely, for which we formulated two hypotheses.

The first hypothesis concerned the content of the programme, i.e. the involvement of real patients. GPTs might consider patient feedback to be of limited use, or might feel vulnerable in asking patients for feedback, which would reflect on their rate of participation.

The second hypothesis concerned the method of application of the programme, which was designed according to a self-directed learning method. This means that GPTs are encouraged to participate in learning activities defined by him or herself rather than defined by a teacher\(^13,14\). GPTs who are aware that they have a lower level of consultation skills, or become aware of this as a result
of the first patient feedback results, might feel the need to develop their skills, which in turn will result in a higher rate of participation in the programme.

In this mixed-method study, we analysed factors that predicted the rate of participation of the GPTs in the patient feedback programme. In a qualitative study, we evaluated whether the rate of participation was hampered by feelings of vulnerability when asking patients for feedback. In a subsequent quantitative study, we analysed the effect of the baseline level of consultation skills of the GPTs on their rate of participation. Analysing the deviant cases is important for the validity of the programme and its results\textsuperscript{15}, and might provide new insight into the requirements needed for effective learning programmes with real-patient involvement.

**Methods**

**Programme and participants**

In 2006, 22 GPTs (two randomly selected classes), in the second trimester of their first-year at the VU medical center in Amsterdam, attended a patient feedback training programme, the purpose of which was to enhance their patient-centred consultation skills\textsuperscript{12}. The training programme fitted into the general training of doctor-patient communication skills, which included educational principles such as ‘feedback’ and ‘self-directed learning’.

Self-directed learning requires formal tools for the self-testing of skills, and for this purpose, we developed and validated a questionnaire, the ‘Patient Feedback questionnaire on Consultation skills’ (PFC)\textsuperscript{16}. During a period of three months, the GPTs were instructed to ask 20 patients to give feedback on their consultation skills, by means of the PFC, which they completed directly after the consultation. Members of the regular educational staff supervised the training programme.
Assessment of baseline characteristics and baseline consultation skills
At the end of the training programme, the GPTs completed an evaluation form containing questions about demographic characteristics (age, gender, clinical experience).

Baseline consultation skills were assessed by observing video-taped consultations of GPTs. The videos were obtained during visits made by standardized simulated patients to the GPT practices before the training programme started. The video-taped consultations were randomly assessed by trained members of the staff of our institute. For the assessment, they used the MAAS-Global instrument, which has a scoring range from 0-6\textsuperscript{17}. A mean score of 3.1 has been set as a standard for the assessment of doctor-patient communication performance\textsuperscript{18}. The simulated patients and the assessors were blinded for the GPT participation rate.

Outcome measures
Qualitative outcome
GPTs who collected less than five PFCs were invited to attend a semi-structured interview with a staff member (HS), lasting approximately 30 minutes, from which we hoped to obtain information about the possible explanations for their lower rate of participation. The tone of the interview was non-accusing, and its purpose was emphasized, i.e. that it was meant to be a learning experience for the staff. The GPTs were asked: a) what reasons they had for collecting so few PFCs, b) what kind of problems they had encountered in organizing patient feedback in their practice, c) whether they were apprehensive about the involvement of patients in their vocational training, d) whether they felt vulnerable or apprehensive about asking patients to give feedback on their skills, and e) whether the communication training, and specifically the patient feedback programme, matched their needs.

Quantitative outcome
The number of completed PFCs that were returned by the GPTs to the research team was considered to reflect their participation rate.
Analysis and statistical methods

A t-test was used to investigate associations between dichotomous baseline characteristics (baseline level of consultation skills, age, gender and clinical experience (all, except gender, were continuous variables) and the rate of participation (univariate analysis). Furthermore, stepwise forward linear regression analysis was performed to examine the combined effect of the baseline characteristics.

Results

Baseline characteristics

Most of the GPTs were female (73%), and in their early thirties (mean age 31.3 years SD 4.0, range 26-45), and with a clinical experience of 2.8 years (SD 2.3, range 0-8). The mean MAAS-Global score was 3.3 (SD 0.7, range 1.8-4.5).

Interviews with GPTs

Six of the 22 GPTs (27%) collected less than five PFCs (range 0-3: 0, 0, 2, 2, 2, respectively 3), and all six were willing to attend the semi-structured interviews. None of these GPTs questioned the potentially beneficial aspects of patient feedback, nor did they question the overall benefits of (scientific) education programmes in general. Although none of them mentioned that they felt apprehensive about asking patients for feedback, half of them had difficulty in selecting patients. The reasons they mentioned for their low rate of participation, and the problems they encountered in organizing patient feedback in their practices, differed widely, and were classified as follows:

- busy time-schedule in general (n=3)
- questionnaire was not considered to be appealing (n=3)
- difficulty in selecting patients (finding appropriate patients, patient refusal [n=3])
- aim of research unclear (introduction course missed [n=3], did not study the manual [n=2], educational aims not clear [n=3])
• resistance against the obligatory character of the patient feedback programme (n=3)
• disappointment about the first results (especially having received socially desirable answers from patients [n=3])
• practice not suitable for organizing patient feedback (n=1).

Participation rate
The mean number of PFCs collected and returned by all GPTs was 10.4 (SD 6.0, range 0-21). Table 1 shows the results of the univariate analysis of the baseline characteristics of GPTs in relation to their rate of participation. The results of stepwise linear regression analysis show that the baseline level of consultation skills (MAAS-Global scores) had no significant effect on the number of PFCs returned (or the rate of participation in the programme). Male gender and more clinical experience tended to be inversely correlated to the rate of participation; the coefficients (B) were 1.061 (SD 0.6), p=.10, and 0.185 (SD 0.1), p=.15, respectively.

<table>
<thead>
<tr>
<th>Baseline characteristics</th>
<th>n</th>
<th>no. of PFCs (SD)</th>
<th>p¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consultation skills (MAAS-Global score)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 3.1²</td>
<td>12</td>
<td>11.0 (2.6)</td>
<td>.71</td>
</tr>
<tr>
<td>&gt; 3.1</td>
<td>10</td>
<td>9.8 (1.8)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 30³</td>
<td>12</td>
<td>12.3 (1.9)</td>
<td>.17</td>
</tr>
<tr>
<td>&gt; 30</td>
<td>10</td>
<td>8.1 (2.3)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>6</td>
<td>6.8 (3.3)</td>
<td>.15</td>
</tr>
<tr>
<td>female</td>
<td>16</td>
<td>11.7 (1.6)</td>
<td></td>
</tr>
<tr>
<td>Clinical work experience (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 2³³</td>
<td>12</td>
<td>11.6 (2.0)</td>
<td>.39</td>
</tr>
<tr>
<td>&gt; 2</td>
<td>10</td>
<td>8.9 (2.2)</td>
<td></td>
</tr>
</tbody>
</table>

¹ t-test for independent samples, ² standard for performance assessment of doctor-patient communication, ³ median value was taken as cut-off point
Discussion

The wide variance in the rate of participation of the GPTs in the patient feedback training programme was used as a basis for further analysis in this study.

The qualitative analysis, which was conducted to provide more information about the reasons for a lower rate of participation, showed that the GPTs had no specific apprehension with regard to patient involvement in their vocational training, or in communication training in general, which we think is reassuring. The reasons for the lower rate of participation differed widely, and were of motivational as well as practical origin.

The subsequent quantitative analysis showed no correlation between the baseline level of GPT consultation skills and their rate of participation in the programme. This means that the GPTs with a lower level of consultation skills did not recognize their deficiencies, which would have increased their motivation to participate in the programme (self-directed learning principle). There was a trend towards a lower rate of participation among male GPTs and GPTs with more clinical experience, the former determinant being in agreement with the results of a study focussing on the attitude of medical students towards communication skills training\(^{19}\).

An interesting reason for a lower rate of participation, which was frequently mentioned by the GPTs in the interviews, was a feeling of resistance against the obligatory character of the programme, without recognizing its obvious benefits. Moreover, if the first results of the patient feedback did not contain any real ‘eye-openers’, some of the GPTs soon became disappointed and gave up. We think that GPTs should be better prepared in how to interpret the results of feedback and take into account the fact that patients tend to give high scores.

The strength of this study is that it was closely monitored and conducted within the setting of an open and stimulating relationship between the teaching staff and the GPTs. This environment makes it possible to try out new educational tools, based on new theories, and to evaluate their expected benefits. However, the drawback of this study is its limited sample size, which
does not allow for further quantitative analysis. It will be a challenge for future research to analyse the effect of the individual rate of participation of GPTs on improvement in their consultation skills. It seems rewarding to make GPTs aware of their baseline patient-centred consultation skills, and the potential benefit patient feedback has in improving these skills. It is likely that close monitoring and coaching of individual GPTs by tutors and the research staff will enhance the feasibility of the project and help GPTs to overcome certain practical problems.

In conclusion, some requirements for self-directed learning with patient feedback have become more clear. This will hopefully enhance the motivation of participants, and, ultimately, benefit the imbedding of a patient-centred approach in general practice education.
References


CHAPTER 8

Reliability of the assessment of consultation skills of general practice trainees with the MAAS-Global instrument

Marcel E. Reinders
Annette H. Blankenstein
Harm W.J. van Marwijk
Dirk L. Knol
Paul Ram
Henriëtte E. van der Horst
Henrica C.W. de Vet

submitted
Abstract

Objectives
The training and assessment of consultation skills is high on the agenda of vocational training institutes for general practice. There is a need for valid and reliable instruments to assess consultation skills in ‘real-life’ settings. We investigated the number of assessors/observations needed for reliable assessments of consultation skills of GPTs with the MAAS-Global instrument.

Methods
Eight teachers at the VU medical center in Amsterdam attended a short training course on the use of the MAAS–Global instrument and assessed the consultation skills of 53 GPTs during 176 randomly allocated video-taped consultations (102 with standardized patients and 74 with real patients). All consultations were assessed twice by two independent teachers. Reliability was estimated with the generalisability coefficient (GC). Mixed model regression analysis was used to demonstrate the effects of ‘individual assessor’ and ‘type of consultation’ on the MAAS-Global scores. The teachers completed a questionnaire to evaluate their perceived competence.

Results
The inter-observer variation was the most important component of variance. An acceptable GC was easier to obtain using real-life patients than using standardized patient consultations: using real patients, for GC>0.8 two assessors and nine consultations were required; using standardized patients, for GC>0.7 three assessors and 30 consultations were needed. The difference in MAAS-Global outcome scores between the individual assessors was statistically significant. Consultations with standardized patients were scored significantly higher than real-patient consultations. The teachers felt sufficiently competent in using the MAAS-Global instrument to assess the consultation skills of the GPTs.
Discussion
Despite the perceived competence of the teachers, to achieve acceptable levels of reliability in the assessment of the consultation skills of GPTs, multiple observations are required, as indicated by the results of this study. We recommend the assessment of real-patient consultations instead of consultations with standardized simulation patients.
Introduction

Adequate communication skills of practicing physicians are related to several important health outcomes, such as emotional health, symptom resolution, function and physiologic measures. There is strong evidence that consultation skills are correlated with patient satisfaction, maintenance of relationships, and therapy compliance. For this reason, much attention is paid to teaching consultation skills to experienced clinicians, as well as in the (general practice) postgraduate training.

The importance of assessment in the process of learning these skills is widely accepted. Assessment of individual progress in the acquisition of consultation skills by general practice trainees (GPTs) during their three years of vocational training in the Netherlands is becoming more systematic, by means of successive observations of video-taped consultations.

The MAAS-Global is the most commonly used instrument to assess consultation skills among the vocational institutes for general practice in the Netherlands. This instrument was developed at the Maastricht University as an educational tool to give feedback to medical students and GPTs. The reliability of this instrument to assess consultation skills of undergraduate medical students has been found to be good, at least under standardized conditions with experienced assessors. Using the generalisability theory, reliability of the assessment with the MAAS-Global instrument was expressed as a function of the number of observed consultations and assessors: two assessors and eight consultations were required for a sufficiently reliable assessment.

What is less known, however, is how reliable assessments with the MAAS-Global instrument are under less controlled and more real-life conditions (e.g., teachers newly trained, and GPTs selecting one of their recently video-taped consultations) that reflect daily practice in a vocational training institute. Individual differences in the scoring of the teachers reduce the objectivity and reliability of the assessments. If there are multiple assessors, taking the mean of the different scores will reduce the inter-observer variability. However, in more widespread use, or in a formative setting (‘education’) rather than in a
summative setting (‘examination’), there will probably not always be sufficient manpower for multiple assessments of the consultations. Furthermore, in contrast to standardized consultations, the video-taped GPT consultations in daily practice will vary in content, technical quality and duration.

This is why we conducted a study to assess the reliability of a test procedure with the MAAS-Global instrument in a setting resembling a real-life GP training institute, i.e. of less standardized conditions. Teachers from the VU medical center attended a short training course in the use of the MAAS-Global instrument. Subsequently, they assessed a mixed set of standardized (simulated) patient consultations and real-patient consultations. Their perceived competence was assessed afterwards.

**Aim of the research**

- To assess the reliability of the test method with the MAAS-Global instrument; and to express the generalisability in terms of the number of observations and assessors needed to achieve reliable assessments.

- To assess the differences in MAAS-Global outcome scores between the teachers and between the types of consultations (standardized versus real-patient consultations).

- To assess the perceived competence of the teachers in the use of the MAAS-Global instrument.

**Methods**

**General design and subjects**

In this observational study, eight teachers from the general practice vocational training staff at the VU medical center in Amsterdam, were trained in the use of the MAAS-Global instrument. This instrument consists of 13 case-independent items, measured on a 7-point scale, referring to communication skills, both in the specific phases of the consultation and in the entire consultation (range 0-6:
0=absent, 6=excellent). Two items were not used: ‘first consultation or follow-up’ and ‘physical examination’. The mean total score was calculated, resulting in a total score ranging from 0 to 6. The standard level of adequate performance of GPs has been set at 3.1 (Angoff procedure and borderline regression)\(^\text{13}\). The teachers subsequently assessed consultation skills by observing video-taped consultations of GPTs. All consultations were scored twice, randomly, by two independent teachers. In the end, all eight teachers completed an evaluative questionnaire about their perception of their competence in assessing GPT consultation skills with the MAAS-Global instrument.

**Educational training**
The training of the teachers in the assessment of consultation skills with the MAAS-Global instrument consisted of: studying the manual; homework assignments (scoring standardized video-taped consultations); and an instruction meeting (during a half-day) supervised by one of the leading experts in this field (PR). The teachers discussed their findings with the expert, and received feedback on their scoring. At the end of the study period the teachers attended an evaluative meeting (2 hours), in which feedback was given to the teachers and the collective and individual results were discussed.

**Materials and allocation scheme**
The video-taped consultations were derived from a patient feedback project, in which 53 first-year GPTs at the VU medical center in Amsterdam were twice visited in their general practices by standardized (simulated) patients, who made video-recordings of the consultations (n=102)\(^\text{14}\). The simulated patients (three professional actors) were trained to play six comparable scenarios of moderate complexity: the (simulated) patients not only had a physical complaint, but also had a ‘second layer’ anxiety or agenda (complete scenarios available on request from the first author). They were randomly assigned to visit each GPT. In the same period, the GPTs provided one of their recent video-recordings of a real-patient consultation (n=74). Both the GPTs and the real patients gave their consent for the video recordings to be used for educational and research purposes.
All consultations were assessed twice. For the first assessment, the 176 videotaped consultations were allocated to the eight teachers (22 each), according to a computerised random-number generator. For the re-assessment, all consultations were re-allocated equally to different assessors, according to a balanced incomplete block design\(^\text{15}\), as is shown in Table 1. Different consultations (cases), with different GPTs, were observed by different assessors (‘fully-nested design’). The teachers were blinded for each others’ scores.

**Table 1. Allocation schemes (Balanced Incomplete Block Design)**

**1a. First 176 assessments**

<table>
<thead>
<tr>
<th>consultations</th>
<th>assessors (teachers)</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A=(1-3)</td>
<td>1A 2A 3A 4A 5A 6A 7A 8A</td>
<td></td>
</tr>
<tr>
<td>B=(4-6)</td>
<td>1B 2B 3B 4B 5B 6B 7B 8B</td>
<td></td>
</tr>
<tr>
<td>C=(7-9)</td>
<td>1C 2C 3C 4C 5C 6C 7C 8C</td>
<td></td>
</tr>
<tr>
<td>D=(10-12)</td>
<td>1D 2D 3D 4D 5D 6D 7D 8D</td>
<td></td>
</tr>
<tr>
<td>E=(13-15)</td>
<td>1E 2E 3E 4E 5E 6E 7E 8E</td>
<td></td>
</tr>
<tr>
<td>F=(16-18)</td>
<td>1F 2F 3F 4F 5F 6F 7F 8F</td>
<td></td>
</tr>
<tr>
<td>G=(19-22)</td>
<td>1G 2G 3G 4G 5G 6G 7G 8G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n=22 n=22 n=22 n=22 n=22 n=22 n=22 n=22 n=176</td>
<td></td>
</tr>
</tbody>
</table>

**1b. Second 176 (re-)assessments**

<table>
<thead>
<tr>
<th>consultations</th>
<th>assessors (teachers)</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A=(1-3)</td>
<td>- 1A 1B 1C 1D 1E 1F 1G</td>
<td></td>
</tr>
<tr>
<td>B=(4-6)</td>
<td>2G - 2A 2B 2C 2D 2E 2F</td>
<td></td>
</tr>
<tr>
<td>C=(7-9)</td>
<td>3F 3G - 3A 3B 3C 3D 3E</td>
<td></td>
</tr>
<tr>
<td>D=(10-12)</td>
<td>4E 4F 4G - 4A 4B 4C 4D</td>
<td></td>
</tr>
<tr>
<td>E=(13-15)</td>
<td>5D 5E 5F 5G - 5A 5B 5C</td>
<td></td>
</tr>
<tr>
<td>F=(16-18)</td>
<td>6C 6D 6E 6F 6G - 6A 6B</td>
<td></td>
</tr>
<tr>
<td>G=(19-22)</td>
<td>7B 7C 7D 7E 7F 7G - 7A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8A 8B 8C 8D 8E 8F 8G -</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n=22 n=22 n=22 n=22 n=22 n=22 n=22 n=22 n=176</td>
<td></td>
</tr>
</tbody>
</table>
Data-analysis

Reliability

The generalisability coefficient (GC) assesses reliability while taking various sources of error (or variance components) into account. This information makes it possible to calculate how many observations and how many assessors are needed to achieve a predefined level of reliability\textsuperscript{16,17}.

We used a mixed-model analysis to calculate the components of variance ($\sigma^2$), by means of ANOVA. The following components of variance were estimated: variance between the quality of consultations (GPT), variance between multiple consultations (video-tape) of GPTs (consult/GPT), and the variance between the assessors assessor (teacher) (assessor/consult/GPT). We calculated the GC with the following generalisability formula (in which $\sigma^2 =$ the variance component, and $n =$ the number of assessors):

$$GC = \frac{\sigma^2_{GPT}}{\sigma^2_{GPT} + \frac{\sigma^2_{consult/GPT}}{n_{consult}} + \frac{\sigma^2_{assessor/consult/GPT}}{n_{consult} \cdot n_{assessor}}}$$

By increasing the number of consultations per GPT or by increasing the number of assessors who judge these consultations, and taking the mean value of these MAAS-Global scores, the variance components are divided by the respective $n$ and the GC becomes higher. High-stakes assessments (summative assessments) generally require a GC of 0.8; otherwise (formative assessments) 0.7 is considered sufficient\textsuperscript{18}. We calculate how many consultations are needed or how many assessors, to obtain an acceptable value for the GC.

Influence of assessor and consultation characteristics

The effect of ‘assessor’ and ‘standardization of consultations’ on the MAAS-Global scores was measured. A mixed-model regression analysis, with ‘assessor’ as random factor, was performed to demonstrate the effect of the individual assessors on the MAAS-Global score (as dependent value), and tested for significance with the likelihood ratio. Univariate linear regression analysis was
applied to test whether the MAAS-Global scores were affected by: the differences in type of consultation (standardized patient versus real patient), baseline characteristics of the teachers (background [GP versus behavioural scientist], gender and work experience of the teacher [in years]).

Evaluation of perceived competence
The evaluation questionnaire contained open format questions concerning the time spent on assessment, and the self perceived competence in the use of the MAAS-Global instrument. The evaluation questionnaire is available on request (MR).

Results

Characteristics of the assessors
The teachers formed a mixed team of general practitioners and behavioural scientists/psychologists (ratio 3/5), gender (male/female: ratio 3/5), and teaching experience (mean 11.3 years (SD 10.4), range 1-32 years). Their mean assessment time per consultation was 29.7 minutes (SD 11.2).

Reliability: aspects of inter-observer variability
When all consultations are taken into account, the calculated scores for the variance components were: GPT (0.08), consultations/GPT (0.36), and assessors/consultations/GPT (0.55). This latter term represents the inter-observer variation.

When the components of variance were estimated for real-patient consultations and standardized consultations separately, it appeared that they were very different. Therefore, in Table 2, the components of variance and the generalisability coefficients were set as a function of the number of consultations and assessors, for real-patient consultations and standardized-patient consultations separately. The variance between the teachers was the greatest source of the total variance, and comparable between the two types of
consultations (0.61 and 0.63, respectively). The GC for the situation when only one consultation is assessed by one assessor (teacher) was 0.22 for the real-patient consultations and 0.05 for the standardized-patient consultations, as can be seen in the upper left corner of Table 2. The GC can be improved by increasing using the mean value of four assessors to 0.41 for the real-patients and to 0.09 for the standardized patients (first row in Table 2). By increasing the number of consultations per GPT (represented in the columns in Table 2) one can see that the value of GC (0.8) is reached by nine consultations per GPT using two assessors, or by 15 consultations when the consults are scored by only one assessor. With three or four assessors, seven consultations are needed. For the standardized patients a GC>0.8 cannot easily be reached: more than four assessors scoring more than 30 consultations per patients are needed. A GC of 0.7 can be reached when three assessors score 30 consultations or four assessors score 25 consultations.

**Effect of ‘assessor’ and ‘standardization of consultations’ on the MAAS-Global scores**

The mean MAAS-Global score for all observations was 3.22 (SD 0.99, range 0.64–5.64), and the mean scores varied among the eight teachers, range 2.59 (SD 0.63)–3.75 (SD 1.41), as is shown in Figure 1.

Mixed-model analysis, with ‘teacher’ as random factor and MAAS-Global score as dependent value, demonstrated the effect of the individual teachers on the MAAS-Global score (Table 3). The likelihood ratio test showed that this random effect was statistically significant ($\chi^2=43.9$ (df=1), $p<.0001$). The background of the teachers (GP or behavioural scientist), gender and work experience as a teacher did not affect the MAAS-Global score (univariate analysis, $p=.97$, $p=.81$ and $p=.11$, respectively), as shown in Table 3. Standardized-patient consultations were related to higher mean MAAS-Global scores (3.51 [SD 0.91]) than were real-patient consultations (2.83 [SD 0.95], $p<.0001$, $t$-test for independent values).
Evaluation of perceived competence

The teachers considered the educational training to be adequate to learn how to use the MAAS-Global (reported by seven teachers, one already had some previous experience). Perceived competence in the assessment of consultation skills was reported by seven teachers (one still experienced some difficulties), which was experienced immediately after the introduction training (by four) or after the first 22 assessments (by two), or unknown (by two). Three teachers specifically appreciated the feeling that they had more ‘grip’ on their assessments, and had developed a more structured approach.

Table 2. Reliability: generalisability coefficients as a function of the number of observations and assessors

<table>
<thead>
<tr>
<th>No. of consultations</th>
<th>No. of assessors</th>
<th>Real-patient consultations</th>
<th>Standardized-patient consultations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>σ^2 (GPT)= 0.22</td>
<td>σ^2 (GPT)= 0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>σ^2 (consultations)= 0.17</td>
<td>σ^2 (consultations)= 0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>σ^2 (assessor)= 0.61</td>
<td>σ^2 (assessor)= 0.63</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.22</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.32</td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.37</td>
<td>0.08</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.41</td>
<td>0.09</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.17</td>
<td>0.13</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0.64</td>
<td>0.18</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0.68</td>
<td>0.21</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>0.71</td>
<td>0.22</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0.32</td>
<td>0.16</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>0.48</td>
<td>0.22</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>0.54</td>
<td>0.26</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>0.58</td>
<td>0.28</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>0.17</td>
<td>0.13</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>0.32</td>
<td>0.18</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>0.41</td>
<td>0.21</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>0.46</td>
<td>0.22</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>0.08</td>
<td>0.16</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>0.18</td>
<td>0.22</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>0.26</td>
<td>0.28</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>0.31</td>
<td>0.28</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>0.28</td>
<td>0.22</td>
</tr>
<tr>
<td>24</td>
<td>4</td>
<td>0.31</td>
<td>0.28</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>0.09</td>
<td>0.15</td>
</tr>
<tr>
<td>26</td>
<td>2</td>
<td>0.13</td>
<td>0.21</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>28</td>
<td>4</td>
<td>0.19</td>
<td>0.28</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>0.08</td>
<td>0.15</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>0.13</td>
<td>0.21</td>
</tr>
<tr>
<td>31</td>
<td>3</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>32</td>
<td>4</td>
<td>0.19</td>
<td>0.28</td>
</tr>
</tbody>
</table>

σ^2 = the variance component; GPT = general practice trainee. Example calculation of a general coefficient (GC) for 2 assessors and 9 consultations for real-patients in the formula:

\[
GC = \frac{\sigma_{GPT}^2}{\sigma_{GPT}^2 + \frac{\sigma_{consult \cdot GPT}^2}{n_{consult}} + \frac{\sigma_{assessor \cdot consult \cdot GPT}^2}{n_{consult} \cdot n_{assessor}}} = \frac{0.22}{0.22 + 0.17/9 + 0.61/(9 \times 2)} = 0.81
\]
Table 3. Results of univariate linear regression analysis to assess the influence of assessor and consultation characteristics on the MAAS-Global scores

<table>
<thead>
<tr>
<th>Teacher (gender, background, working experience in years)</th>
<th>MAAS-Global score Mean (SD)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (F, MSc, 1)</td>
<td>2.96 (0.62)</td>
<td>&lt;.0001¹</td>
</tr>
<tr>
<td>2 (M, MD, 19)</td>
<td>2.89 (0.91)</td>
<td></td>
</tr>
<tr>
<td>3 (F, MD, 15)</td>
<td>3.75 (1.41)</td>
<td></td>
</tr>
<tr>
<td>4 (F, MSc, 2)</td>
<td>3.74 (0.79)</td>
<td></td>
</tr>
<tr>
<td>5 (M, MSc, 32)</td>
<td>3.51 (0.91)</td>
<td></td>
</tr>
<tr>
<td>6 (F, MSc, 7)</td>
<td>2.59 (0.63)</td>
<td></td>
</tr>
<tr>
<td>7 (M, MSc, 8)</td>
<td>3.26 (0.72)</td>
<td></td>
</tr>
<tr>
<td>8 (F, MD, 6)</td>
<td>3.68 (1.06)</td>
<td></td>
</tr>
</tbody>
</table>

Teacher characteristics

| GP (n = 3) | 3.24 (1.19) | .81 |
| MSc (n = 5) | 3.21 (0.84) |      |

| Male (n = 3) | 3.22 (0.88) | .96 |
| Female (n = 5) | 3.23 (1.05) |      |

Work experience (in years) | .11 |

Acquired experience

| Assessments (n = 176) | 3.18 (1.03) | .38 |
| Re-assessments (n = 176) | 3.27 (0.94) |      |

Consultation

| Standardized-patient | 3.51 (0.91) | <.0001¹ |
| Real-patient         | 2.83 (0.95) |          |

¹Likelihood ratio test; F=female, M=male; MD=Medical Doctor; MSc=Master of Science (psychology, behavioural sciences)
Figure 1. Box and whisker plot of mean MAAS-Global scores per teacher

Boxes represent the 25, 50 (median) and 75 cumulative relative frequencies (percentiles); whiskers represent the range; dots represent outliers
Discussion

Main findings
The teachers felt competent in using the MAAS-Global instrument to assess the consultation skills of GPTs, even in the early stage of the training, and generally valued its educational potential. However, the inter-observer variation was high: the assessor was a significant factor of variance. Our analysis by means of the generalisability theory demonstrated that the number of consultations (rather than the number of assessors) needed for reliable assessment of the consultation skills of GPTs greatly depended on the type of consultation. Real-patient consultations received significantly lower MAAS-Global scores than standardized-patient consultations, but also showed more variance (0.22 versus 0.05), and therefore less real-patient consultations are needed to be scored for reliable assessments, compared to the standardized-patient consultations. Apparently the quality of the consultation of the GPT differed more among the consultations of the real patients than among the consultations of the standardized patients.

Interpretation
The difference in scoring among the teachers was statistically significant but could not be explained by the different educational background, gender, or work experience of the teachers, because these variables did not seem to correlate with the MAAS-Global scores. However, the number of assessors (8) was too small to allow conclusions. It is remarkable that, when asked, the majority of them were aware of their own tendency to be critical or mild in their appraisals.

The type of consultation (standardized consultations versus real-patient consultations) influenced the height and variation of the outcome score. This might be due to the fact that real-patient consultations are more influenced by a variety in factors such as the length of the consultations, the technical quality of the video-tapes of the consultations, or their medical content (‘sample criteria’). It was surprising to know that real-patient consultations had a higher reliability than standardized patients.
Comparing the results of the performance of the MAAS-Global instrument with other studies is hampered by a difference in standardization of the scoring conditions, for example consultations in daily practice or multiple-station examination\textsuperscript{12,19,20}. It is important to realize that the reliability of the assessments of consultation skills depends on the intrinsic qualities of the MAAS-Global instrument as well as on the test-procedure ('setting'). It is the latter aspect we focused upon here.

It is likely that the inter-observer variability might be reduced by more intensive training of the assessors. In a previous study it was found that for reliable assessment of the consultation skills of experienced GPs with the MAAS-Global instrument (GC 0.8), two assessors and eight (standardized) consultations were needed\textsuperscript{12}.

**Strength and limitations**
Our study, with teachers from a general practice vocational training staff, reflects the reality of daily practice in the assessment of consultation skills. The teachers experienced the intricate nature of scoring consultations. The assessors were closely monitored, which contributed to the fact that we only had a minimum of missing data.

Our study design included a large number of observations, which were all scored twice, by different assessors, in a randomization scheme that provided a maximum of variability.

Unfortunately, our study design in which video-tapes were assessed and re-assessed did not make it possible to give interim feedback to the teachers on their individual assessment performance (for example by means of benchmarking). It is likely that such interim reinforcement feedback, aimed at self-correction, might have a beneficial effect on reliability.

**Practice implications and future research**
What can we do to optimize the generalisability of assessing the consultation skills of GPTs? Apparently, assessments of real-patient consultations lead to a higher GC than assessments of standardized consultations do.
Demands will be made on the acquired experience of the teachers in similar future projects. It is interesting to examine whether feedback on individual performance of teachers ('hawks' and 'doves') will confirm or cause adjusted behaviour in the future assessment of consultations. If the individual assessment styles of teachers are already known, in future settings it would probably be wise to make matched pairs of critical and mild assessors (at both ends of the scale), which would then balance out their differences.

**Conclusion**

In conclusion, despite the feelings of competence that the teachers experienced, acquiring acceptable levels of reliability in the assessment of consultation skills remains difficult. The consequences for practical use in a real-life setting in the vocational training institute for GPs are that multiple observations are needed, as indicated by the results of this study.
References


CHAPTER 9

General discussion
How should we valuate the patient feedback programme? What did the GP trainees learn from patient feedback, and did they appreciate it? Due to the complexity of an educational intervention in terms of effectiveness, impact, learning potential and appreciation, it is not possible to draw simple conclusions. In this general discussion we once more reflect on the above questions. All phases of the research project will be addressed: from the feasibility study (in which needs were assessed), to the study of the validity of instruments, to the effectiveness study.

Before going into a detailed description of various items, a summary of the key findings is presented here.

- In the literature, evidence of the effectiveness of feedback assessments from patients is scarce, and is hampered by the heterogeneity of study designs and the lack of rigorously validated instruments (Chapter 6).
- There is evidence that real patient feedback interventions are effective on lower levels of educational impact, but not on the level of assessment of actual change (improvement) of the consultation skills of physicians (Chapter 6).
- Patient feedback is considered to be a major potential for acquiring consultation skills, but it is limited by several organisational obstacles, and by socially desirable answers from patients (Chapter 2).
- Younger patients and male patients, the language proficiency of the GP trainee, and the complexity of the consultation, are all correlated with lower patient feedback assessments (Chapter 4).
- Specific components of the patient-centeredness of communication, such as ‘finding common ground’ and ‘personal context’ were more critically evaluated by patients (and GPTs) than ‘exploring the disease’ (Chapter 4).
- The Patient Feedback questionnaire on Consultation skills (PFC) appears to be a valid, internally consistent instrument, it adequately covers the corresponding ‘communicator’ competency items, and it has an acceptable scoring scale (Chapter 3).
• The patient feedback training programme did not improve the consultation skills any more than the regular doctor-patient communication education (Chapter 5).

• A sub-group of GP trainees who actively participated in the patient feedback training programme did more improve their consultation skills than GP trainees who did not actively participate (Chapter 5).

• The rate of participation of GP trainees in the (self-directed learning) patient feedback programme varied widely. However, this was not related to their baseline level of consultation skills or feelings of vulnerability/awkwardness when asking patients for feedback (Chapter 7).

• The assessment between teachers of consultations skills of GPTs varied significantly. For a higher reliability of the assessments, estimated with the generalisability coefficient, the use of real-patient consultations required less observations than observations of standardized simulated patients (Chapter 8).

These key findings have, in turn, given rise to many new questions. What could be the reasons for the limited overall effect of the intervention on GP trainees? Was it the methods we used, or the clinimetric qualities of the instruments, and did we inform, stimulate and monitor GP trainees enough to make them actively participate? Did we adequately check the participants’ needs?

This general discussion presents an overview of topics that refer to the possible role of patient feedback as a potential learning method. References are made to the corresponding chapters in this thesis, or to the literature. For better understanding of this chapter, the sub-headings cover sections which are all structured to include: an introductory paragraph; a brief summary of findings; a discussion of the findings; and practice implications for future education and/or research.

In the following chapters, a personal selection of some important subjects will be discussed that concern: research on patient feedback in general practice, learning aspects in general practice training, and research on medical education.
Chapter 9

Research on patient feedback in general practice training

Assessment of needs
In order to make successful renewals or changes in the educational programmes in general practice training it is of great importance to carefully assess the needs of all participants: the staff of the vocational training institute (locally or nationally), the research staff, and the GP trainees. Research on educational interventions should also start with adequate needs assessment, which increases the adherence of participants to the intervention. Needs assessment can produce conflicting results, for example when the research staff suggest a more intensive and time-consuming programme than the educational staff or the GP trainees consider being feasible. From the palette of currently available methods for the assessments of needs, we applied personal interviewing, evaluation forms, and discussion groups with GP trainees and their trainers. In the specific field of research on patient involvement in general practice education, personal interviews were held with patients and representatives of a patient organisation.

Summary of findings
In the exploratory study (Chapter 2), the results of semi-structured interviews and discussion groups showed that the GP trainees and their trainers were of the opinion that patient feedback has a great potential to improve consultation skills. Most of the GP trainees expected that patient feedback would provide enough educational material for discussions with their GP trainers, without apprehensiveness about the ‘extra’ assessment or feelings of vulnerability. Although they were eager to know what the patients actually thought about their consultation skills, some feared unreasonable criticism. The GP trainees also mentioned that they already had a full agenda, and that a project like this would certainly be an additional burden. The evaluation forms of GP trainees showed that acquiring consultation skills was high on the learning agenda (next to clinical skills) for the GP trainees.

The patients were generally positive to very positive about the idea of giving feedback to the GP trainees. They expected patient feedback to benefit
both doctor and patient, and were generally glad to make a contribution to the GP training programme. The concept that their feedback would be used to improve the consultation skills of GP trainees was clear for most of the patients. Only a few patients admitted that they found it difficult to give negative feedback to a doctor (especially if it was not anonymous).

Discussion
Careful introduction of a newly introduced educational programme, such as patient feedback, seems to be more important when it includes sensitive assessments of personal skills. Although, fortunately, the participants seemed to accept the concept of patient feedback, both the research staff and the educational staff feared that the GP trainees and patients would be apprehensive about participation. Vulnerability in being assessed, the unequal balance of power between doctor and patient, fear of the consequences etc., there are many reasons why the patient feedback programme might encounter obstacles. However, we seemed to have been able to match the needs of the participants to the content of the intervention, which has great impact on the validity of the programme and adherence to the programme.

Practice implications
The importance of assessing the needs of all participants can easily be underestimated. In future research on the implementation of educational programmes, we might even more deliberately ask ourselves the following questions: Have we really checked the patients’ needs? And have we checked on the GP trainees’ needs? Does the programme fit into their schedule? And what if these answers are conflicting, have we convinced them of the true benefits of educational research and the implementation of new education programmes?

Concept and design
Asking patients for oral feedback on the quality of the consultation at the end of a consultation is quite relevant, but students (and teachers) are looking for more general, aggregated or context-independent forms of feedback and skills assessment. As the systematic review (Chapter 6) demonstrated, almost all of
the patient feedback studies that were included made use of questionnaires for the assessment of skills\textsuperscript{1-14}. Questionnaires make it possible to study patterns and comparisons. However, there seems to be no preferential method with regard to how the feedback assessments (results of the questionnaires) should be transferred to the receiver. Personalised (tailored) reports or coaching sessions in which the collated summary of feedback is discussed are most frequently used. The intensity of patient feedback programmes, as described in the literature, varies widely with regard not only to the number of patients per physician, but also the duration of the intervention. Discussing the results and the meaning of patient feedback with a GP trainer, a teacher, or among peers, is invented to further intensify the learning effect.

Summary of findings

From the feasibility study (Chapters 2 and 4) we concluded that the educational value of patient feedback was correlated to the context of the consultation. Apparently, when GP trainees had to deal with relatively simple medical issues, patient feedback produced less meaningful learning material. In view of the common finding that first-year GP trainees see a high percentage of patients with relatively simple medical problems, they were advised to select those patients from whom they expected to receive relevant feedback. We therefore chose a method for acquiring patient feedback in which the GP trainees themselves select patients and ask them for feedback on their consultation skills, directly after the consultations. Based on the literature, the number of patients per GP trainee was set at 30, but in the following trial it was set at 20, because the GP trainees considered that acquiring feedback from 30 patients was an extra burden on an already busy time-schedule. The length of time needed to acquire this number of patients was set at three months.

The study material that was provided with the patient feedback programme included assignments for GP trainees, in which they discussed the collated results of patient feedback with their trainers. Approximately one third of the GP trainees completed the assignments, one third partly completed them, and one third did not complete them at all.
Discussion
The methods of patient feedback described in the literature are too diverse to make it possible to draw any conclusions about which is the best. The intensity of the programme is not always clearly correlated with the educational impact, and choices are also mainly based on local and cultural customs.

We expected the GP trainees to receive more valuable feedback if they themselves selected the most appropriate patients. This method is new and unique, and was expected to produce substantial educational impact. Experiential practice-based learning, in which the GP trainees themselves select patients to learn from, fits into the generally applied learning methods that are already used in the vocational training provided by the VU medical center.

Practice implications
Obviously, the intensity and duration of patient feedback should be sufficient to produce actual effect or change. It is arguable that the impact of our method could gain in strength. The choice of intensity should be a sensible one: a programme that last for many years, or require the involvement of a great many patients, or inflicts increased costs or a burden on the system, will defeat its own aim.

A possible strategy that can be applied to gain more strength is to make patient feedback a more prominent topic in the learning discussions with the GP trainer, in which case the trainers should be prepared to participate more actively. Nowadays, the agendas of these discussions very much depend on the input of the GP trainees, but they could be made more structured and test-driven. A structural and prolonged focus on the perspective of the patient can be combined with the perspectives of the GP trainers or the co-workers (360 degrees feedback), which could enhance the educational effect, at least under conditions in which skilled facilitators encourage reflection and set concrete goals\textsuperscript{15}.

The patient sample selected by the GP trainees is not a random sample, so the patient feedback assessments can not be considered to be objective outcome measures. If there is need for objective patient feedback, clear assessment of the patient’s characteristics and the content of the consultation is
a prerequisite. Questionnaires should then be handed out to patients at random by a research assistant or a practice assistant. This may be a feasible alternative, especially when GP trainees appear to be struggling to complete their assignment, but of course this excludes the particular advantage of receiving feedback after specifically valuable consultations. Again, this method, in which there is no random selection of patients, precludes setting a standard, and does not allow the benchmarking of scores in a cohort of GP trainees.

The PFC questionnaire
The questionnaire we needed had to cover the whole domain of general consultation skills, the principles of patient-centeredness, and the meaning and content of the key competency ‘communicator’. Furthermore, it had to be practical and not to be much of a burden to complete for the patients and the GP trainees, and it had to be relevant and appealing. In general, it ought to be the right learning instrument for the development of consultation skills, and ultimately help GP trainees to reflect on their consultation skills and guide discussions with their trainers or their patients.

Many of the existing patient satisfaction questionnaires share a drawback: ceiling effects\textsuperscript{7,16,17}. This scoring on the upper end of an answering scale hampers the possibility to measure change (because there is little room for improvement), and also limits the learning impact.

Summary of findings
We began the validation process with, as a basis, an already existing questionnaire, the PPPC (patient perception of patient-centeredness), because this questionnaire already covered a substantial amount of general consultation skills (Chapter 3). A further advantage is that it consists of a patient version and a doctor’s (self-assessment) version, which challenges the participants to reflect on their performance and the learning possibilities. In order to cover all items of the key competency ‘communicator’\textsuperscript{18}, we extended the PPPC to include items controlled for face validity (by involving patients and GP trainees) and content validity (by involving a panel of experts).
The resulting questionnaire, called the Patient Feedback Questionnaire on consultation skills (PFC), contains 16 questions (on a single page) and takes only a few minutes to complete\textsuperscript{19}. The technical qualities seem to be satisfying: factor analysis showed a one-dimensional construct with a high internal consistency was high (Cronbach’s alpha 0.89). For the single items, the mean item-response rate was 97.3\%, and the mean percentage of maximum scores (ceiling effect) was 67.5\%. In conclusion, the PFC appears to be a valid, internally consistent instrument.

Discussion

The PFC is based on a successful combination of two principles: patient-centeredness and the key competency ‘communicator’, which apparently share a comparable concept (one-dimensional construct). It is reassuring that other recent studies have also developed validated questionnaires\textsuperscript{17}, with comparable content (eye-ball test), which reciprocally enhance the external validity of both questionnaires.

The efforts we made to prevent ceiling effects, by selecting and rephrasing questions, has produced acceptable results. However, ceiling effects in the scores of patient (satisfaction) questionnaires are difficult to overcome. Socially desirable answers, as a result of inequality in the balance of power between patients and doctors, might be reduced by improving patient education and information. Moreover, patients might really be satisfied with the communicative aspects of a consultation with a GP trainee, which can be considered as a major quality of GP trainees. However, from the perspective of the teaching staff, who regularly assesses video-taped consultations, there is still considerable room for improvement in many cases.

Practice implications

The PFC is a useful, practical and attractive instrument for transferring feedback from the patient to the GP trainee. We do not claim that this is the ‘final version’ of the PFC, because the validation process is dynamic. Further external validation might indicate the need for modifications and improvements.
Although patients tend to answer on the upper end of the scale, it is the interpretation of these answers, in particular, that gives the impression of limited learning potential. GP trainees tend to interpret the results as too flattering. We need to convince them that only excellent results are satisfactory. Some favour the idea of dichotomising the outcome into ‘excellent’ ratings (maximum scores) versus all other ratings, which visualises the deficiencies more clearly and is therefore more educative.

Training and organization
Adequate estimation of time-burden, financial costs, availability of staff, and other feasibility aspects, is essential for such complex projects, otherwise they will lose their impact. The patient feedback training programme was intended to prepare, educate, and ultimately facilitate GP trainees to work with patient feedback and to benefit from it.

Before the GP trainees started asking patients for feedback they first attended a training session, in which they were informed about the content and meaning of the programme. All course materials were both provided digitally and in print form. The GP trainees could practise, with a simulated patient, how to ask a patient for feedback. They also discussed organisational aspects, and the pitfalls of patient feedback in daily practice. The training was supervised by the educational staff, who received instructions (in a two-hour meeting), a protocol, and a checklist of the items to be discussed with the participants.

Summary of findings
Despite the comprehensive preparation, according to the GP trainees, obtaining patient feedback in their practice was not an easy thing to do. The majority of the GP trainees (67.4%) encountered various obstacles in organising the programme in their practice that hampered the inclusion of patients for feedback (Chapters 2 and 7). In general, these obstacles were related to GP trainees not being well informed, a busy time-schedule, or basic organisatorial problems.

The progress of participation was monitored by the teaching staff. At regular intervals the GP trainees were given encouragement in an open, ‘non-authoritative’ way (by letter or e-mail), and were asked about their progress in the
project. The research team was not involved in motivation to adhere to the educational programme or individual progress.

**Discussion**

The importance of a well-designed and well-organised training programme may seem obvious. It is likely to be related to the motivation of the GP trainees to actively participate in the programme, at group or individual level. However, in view of the obstacles mentioned by GP trainees it is questionable whether they were sufficiently informed and facilitated. It is always a delicate balance between providing enough information and overdoing things (being intrusive). Close monitoring, based on individual needs and motivation, is probably the best approach.

**Practice implications**

The literature on patient feedback intervention studies often lacks adequate descriptions of process evaluation with regard to the various aspects of training and organisation. In our opinion, adequate descriptions are of major importance, enabling new projects in general practice or other medical specialties, to benefit from them, by learning from the pitfalls, and enabling them to reproduce or improve on the results. In future projects of this kind, the participation of GP trainees could be enhanced by checking individually what is needed before they actually start implementation in their practice, and by monitoring the experiential learning process closely.

**Standardized simulated patients**

The assessment of consultation skills is related to the context of a consultation, therefore standardized (simulated) patients can be needed. Doctors are usually aware of the fact that they are being observed or assessed (unless incognito patients are involved). The performance of consultation skills of GP trainees is therefore usually assessed in the daily professional practice, which is their natural habitat. This approaches better the performance of GP trainees ('does-level'), instead of in a controlled, standardized setting in which the GP trainees can demonstrate what they can do ('shows-how-level'). An alternative method
that is well described\textsuperscript{21} is prolonged video-recording of all subsequent consultations within daily practice. A selection of video-taped consultations against sample-criteria can than be made.

\textit{Summary of findings}
In our project, three simulated patients (professional actors), who were trained to act in six consultation scenarios of moderate and comparable complexity, visited all practices. The visits were announced, and the GP trainees gave their consent before participating. All consultations (one scenario at each visit) were video-taped, so that they could be assessed by the teaching staff (Chapters 5 and 8). The simulated patients encouraged the GP trainees to behave naturally, and not to put more effort into the consultation or ask more questions than they would normally do. The simulated patients were also trained to give individual feedback to the GP trainees after the visits.

\textit{Discussion}
With incognito standardized patients the GP trainees would be unaware that they are being assessed. We chose not to use this method, mainly because in such projects implementation of the fielding procedure is quite difficult to realize, and also because of the excitement that such projects cause among GP trainees\textsuperscript{22}.

Well-trained professional actors are necessary in such projects. They have to portray a credible patient and a credible consultation, and do so repeatedly under the same conditions, so that the GP trainees are facilitated in behaving naturally in their role as a doctor. Furthermore, more than half of the teachers who finally assessed the video-taped consultations mentioned that the professionalism of the actors ‘coloured’ their scores (Chapter 8).

\textit{Practice implications}
We recommend the observation of consultations with standardized patients in the general practice location, even though the training and involvement of simulated patients is time-consuming (and costly). It is a method that reflects most clearly how GP trainees actually perform under comparable conditions,
although the assessments are still on the ‘shows-how’ level. If assessments on the ‘does’ level are preferred, labour-intensive and fragile procedures, such as fielding incognito procedures, are needed. The method of prolonged video-recordings in daily practice and inclusion of consultations against sample criteria seems to be a valid, though costly and strenuous, alternative.

**Building up evidence of effectiveness**

The application of new educational methods should be guided by scientific evidence of their effectiveness. There is little evidence of the actual learning potential as a result of patient feedback in the postgraduate training of young physicians or the professional training of more experienced physicians. Furthermore, the level of impact of most of the earlier studies is limited to ‘appreciation and valuation’ data (‘happiness data’). One can imagine that the contribution of these studies to actual scientific proof or evidence is limited, and depends much on the rigorousness of the methods used to acquire qualitative data. Actual change in behaviour as a result of the introduction of patient feedback is much more difficult to demonstrate, possibly due to the lack of precision in outcome measures, or the lack of responsiveness of the measurement instruments.

**Summary of findings**

As demonstrated in the review (Chapter 6), there are several levels of educational impact as a result of patient feedback, and several corresponding levels of evidence of effectiveness. The chapters in this thesis describe a diversity of consecutive studies (from qualitative research to a controlled trial), that all build up integrated evidence of the effect of patient feedback (interventions).

There is less evidence for effectiveness of these studies at higher levels of educational impact. The GP trainees seemed to valuate/appreciate patient feedback, and they frequently formulated learning topics as a result of the feedback they received (Chapters 2 and 6). At the highest level of educational impact, in the trial (Chapter 5), we found no direct evidence of the effectiveness of patient feedback (intention-to-treat analysis); it seems only to be effective for
a sub-group of GP trainees who actively adhere to the patient feedback programme.

Discussion

Acquiring evidence of effectiveness at the highest level of educational impact (actual change) can be difficult, due to the inherent complexity of interventions, the usually small number of participants involved, the lack of precision of outcomes measures, and the limited responsiveness of instruments with which the outcome measures are assessed\(^\text{23}\). Our intervention seems to be no exception.

Consultation skills cover a large variety of qualities. One can easily imagine that assessing improvements in consultation skills in general is far more difficult than assessing only one single skill or quality. The MAAS-Global instrument that we used to measure general consultation skills mainly covers the same areas of the consultation as the PFV, but it has previously been criticised for lack of responsiveness\(^\text{24}\). However, in our trial, a mean improvement in the consultation skills of all GP trainees (intervention group and control group) was found during the early stage of their vocational training.

Practice implications

An important remaining is: what level of evidence or support is needed before including new learning opportunities, such as patient feedback, in a GP vocational training programme is warranted. In the past, the answer to that seems to depend on personal and local preferences for teaching (patient-centered) communication skills. In future, we recommend awareness of the level of acquired educational evidence and the consequences this has for incorporation new learning methods.
Learning in general practice training

The definitions of the required consultation skills for GP graduates are described in the competency profiles\textsuperscript{18}, so it is important for teachers to know how GP trainees can reach this end-goal efficiently. What is the most efficient and effective learning method? Apparently, not all GP trainees have the same learning needs, and especially not all at the same stage in their training. Therefore, in recent years at the VU medical center, acquiring knowledge has become more characterised by the self-directed learning method, in which not the teacher, but the student sets the learning agenda\textsuperscript{25}. GP trainees are encouraged to plan ahead with regard to the topics they want to learn, and discuss these with their teachers and peers. The patient feedback programme, with emphasis on practice-based experiential learning, was designed to fit into the general training of consultation skills. Two aspects that have strong educational impact on the performance of practicing physicians are 1) performance assessments\textsuperscript{26}, and 2) feedback\textsuperscript{27}, both of which will be discussed below.

Assessments of GP trainees
Assessments of GP trainees are considered to be important, not only for measuring the progress of clinical performance (summative assessment), but also because they have a strong educational potential (formative assessment). In the Dutch new curriculum of the vocational training for GPs, the GP trainee’s performance with regard to consultation skills is assessed regularly. Unfortunately, there is insufficient evidence of the reliability of these assessments.

Summary of findings
In the trial study (Chapter 5), 176 video-taped GP trainee consultations were observed by the trained staff of the institute for general practice to assess the consultation skills of the GP trainees with the MAAS-Global instrument, which has been tested for validity and reliability. With the generalisability theory\textsuperscript{28,29},
the number of consultations per GP trainee and the number of assessors per consultation could be calculated for achievement of reliable assessments (Chapter 8). The difference in scoring between the assessors was significant. The assessors appreciated working with the MAAS-Global instrument as they gained experience and competence.

Discussion
For summative assessments, the number of consultations needed to obtain reliable assessments was high, but greatly depended on the type of consultation that was observed, whether it were standardised consultations with simulation patients, or consultations with real-patients. For a higher reliability of the assessments, we recommend the use of real-patient consultations, of which less observations are required than observations of standardized simulated patients. For the teachers, the formative assessments of the consultations, and possibilities to give feedback to the GP trainees, probably have greater meaning.

Practice implications
At our institute, we now have a group of teachers who are experienced in the use of the MAAS-Global instrument. The results of their mean scores were fed back to them. Awareness of being critical or lenient might affect their future scorings or reduce the inter-observer variation. It might be interesting to make combinations of known critical assessors and known non-critical assessors ('hawks' and 'doves'), for a reciprocal averaging effect.

Feedback
The effect of feedback on practicing physicians has been widely studied in recent years. It appears that feedback is one of the more powerful methods with which to stimulate learning in students. The Pendleton rules provide certain conditions for safe and effective feedback, which should be conducted respectfully, non-judgementally and in a friendly way, and also be constructive (with the intention of achieving improvement). Furthermore, it should be bilateral, and leave options for agreements about intentions as to how to continue. Moreover, timing and dosage are also important aspects.
**Summary of findings**

Apparently, GP trainees are far more interested in criticism than in positive feedback. An important finding in our studies was that the GP trainees indicated that too much positive feedback decreased their motivation to learn, because they considered it unrealistic. They stopped handing out questionnaires, which further limited the learning effect. In Chapter 4, we studied the differences in opinions about the quality of consultation skills between patients and doctors, and found that GP trainees were far more critical in their self-assessments than the patients were in their feedback assessments, although they also showed similarities in the pattern of the different components of patient-centeredness.

Feeling vulnerable when asking patients for feedback, anxiety for the results of the feedback assessments, a non-comfortable learning environment, and apprehensiveness about disturbing the patient-doctor relationship might seem to be obvious barriers for participation\(^3\), but they were seldom mentioned by the GP trainees.

**Discussion**

Considering whether feedback from patients by means of a questionnaire is given according to the conditions for feedback as described above, we think it generally has the following the qualities: it is respectful and friendly, non-judgemental and constructive, aiming at improvement. What is missing is a direct contact with the patient, and the fact that it is not bilateral and therefore leaves no room for agreements about intentions and plans on how to continue. Moreover, we do not know for sure whether the GP trainees felt that it was safe to ask patients for feedback, despite the fact that they denied feelings of vulnerability. From the perspective of the patients, it is unsure if they were convinced that critical and honest feedback could benefit the development of GP trainee’s consultation skills, and therefore withhold them of potential feedback material.

**Practice implications**

Neither the patients nor the GP trainees, like most other people, are (constantly) aware of their communication behaviour, and they are not trained to register or
rate their skills. This takes time and critical reflection. More intense and personalised education and information for patients might solve this problem. We should also prepare each participating GP trainee individually of the type of answers they can expected from patients and how to handle them. It seems a paradox that critical feedback from patients, which might reflect their dissatisfaction, could also be considered as a high learning potential (formative setting). An open and equal relationship between patient and doctor is then a prerequisite.

**Learning incentives from patient feedback?**
Feedback from patients provides insight from a unique perspective. As in other societal sectors in which policies are made after thorough marketing research, client satisfaction can lead to better service and better quality of the care that is provided. More understanding of how and why GP trainees can learn from this new perspective is desirable. GP trainees should be aware of their personal qualities and deficiencies, and should recognise what feedback can offer them. This process has to be monitored by teachers or GP trainers if it is to succeed.

**Summary of findings**
The patient feedback programme is based on experiential learning. This implies that along the way, the GP trainees were supposed to learn about their qualities and deficiencies from their patients’ perspective. For example, in general, the patients were quite satisfied with the diagnostic phase of a consultation, whereas they were less satisfied with finding common ground or the personal context of the consultation. Because the feedback was also consultation-specific, reflection on what just had happened during the consultation provided all kinds of possible learning potentials.

To some extent, most of the patients seemed to be able to give feedback. We do not know how well patients were being informed and motivated by the GP trainees to give feedback, and finding patients who can make a real contribution is therefore a challenge for the participating GP trainees.
Discussion

It is likely that not all GP trainees are suited for receiving patient feedback, because it probably requires a certain state of mind, a certain sensitivity or susceptibility. The following example may demonstrate this. A patient reports: 'if the doctor had the right patient file in front of him, I would have had far more confidence in him'. Although this feedback seems to offer major learning potential, in the self-assessment form the GP trainee only mentioned ‘an overly critical patient’. The question is: do first-year GP trainees already have these qualities, or are they pre-occupied with so many other relevant things (e.g. making medical diagnoses)?

Giving feedback to the GP trainees on their baseline level of consultation skills (and optionally, benchmark scores), before the start of the patient feedback programme could probably motivate the GP trainees to actively participate and learn. To ensure random assessment of pre-intervention and post-intervention measurements in our trial, we did not do so.

Practice implications

It is interesting to identify the characteristics of GP trainees who benefit (or do not benefit) from a patient feedback intervention, or from educational interventions in general. Do these GP trainees have different demographics, different learning styles, different baseline skills, or do they have different needs?

It may be interesting to offer this programme to more experienced GP trainees (in their third year, the year in which there is more contemplation). However, from the results of the literature review (Chapter 6), there seems to be no correlation between the educational stage (student or experienced physician) and the intention or ability to change.

Projects like this need close monitoring by the teaching staff and GP trainers. It is important to know what motivational teaching actually can contribute. We need to find ways to measure the content and impact of discussing the learning topics, as a result of patient feedback, or in general.
Research on medical education

In addition to offering new insights into learning from patients, the studies in this thesis have demonstrated the importance of close collaboration between practicing GPs and/or GP trainees, the staff of the vocational institute, and researchers. Ideas for research topics should be shared between the three parties, discussed with regard to relevance and practicality, and agreed upon. Ideas that do not meet these requirements are deemed to fail. Figure 1 presents a diagram of this co-operation.

Figure 1. Research on medical education
Has the patient feedback project met these criteria? The idea of including patient feedback in the vocational training in order to give patients more say, originated from the vocational training staff and the research staff. The proposed participants, the GP trainees and the GP trainers were informed in good time, and open discussions provided ideas for improving the programme. A year before the actual trial started, a feasibility study was conducted, to optimise the effectiveness of the programme, and to overcome pitfalls (Chapter 2). In this respect, we can conclude that the basic requirements were met.

This thesis has shown that the amount of effort that is needed to make educational interventions ‘work’ should not be underestimated. Although the intention is to meet the needs of all participants, it should be kept in mind that realistic goals must be set. Health workers might be reluctant to value research on the effectiveness of educational interventions, because of different needs and different views about ‘what works’\(^\text{23}\). Furthermore, it is probable that not all GP trainees will appreciate the patient feedback programme, and it is most likely that not all GP trainees will benefit from it.

It should be kept in mind that the intervention is ‘work in progress’: the patient feedback programme and the implementation procedure are new, teachers and GP trainers are still getting used to it, and it is becoming more tailored to the needs of the GP trainees. It might even take several years before the programmes will be integrated in the postgraduate training.

How can we do better? From the experience of the Centre for Studies in Family Medicine, London, Canada, we learned that for a period of at least three months researchers must be present in classes at the time of implementation of a new intervention in order to get it off the ground. You cannot simply deliver a research project at the door of the institute and expect it be carried out as it was meant to.

The educational training of GP trainees tends to be more practice-based, and less centrally planned. We did not go to the practices to monitor the learning moments and the individual progress of the GP trainees, because we considered that to be too intrusive. We think that it might even reduce the effectiveness of patient feedback.
Research on communication skills is probably the most common topic in research on education in primary care. Communication education represents approximately 25% (high share) of the education time at the institute during the first year of the GP training. The studies described in this thesis and other studies\textsuperscript{2,7,16,35} have shown that the communication skills of GPs and GP trainees are rewarded with high patient satisfaction rates. To maintain the relatively high standard of communication, it is justifiable to continue with education and research on communication as is the standard at the moment. However, it also gives rise to the question of whether or not research on other key competencies should receive more attention. One can think of collaboration, professionalism, etc. Irrespective of the specific field that is studied, for educational research in primary care it remains important to know how GP trainees learn and why they learn.

Research questions and study designs should be realistic, and have a reasonable expectancy to measure a change in outcomes. It seems that higher levels of evidence require high levels of effort. Scientific proof of effectiveness can be achieved at various levels, from valuation of the project to an actual assessment of change. However, although it is costly and time-consuming, major scientific proof should be acquired by designing educational interventions that apply rigorous standards of reliability and validity, whether they are (randomised) trials or qualitative studies\textsuperscript{36}. As an example, we refer to the vast number of patient satisfaction studies, and, in sharp contrast, the limited number of studies that measure their learning potential.

This thesis has shown that carrying out research in the field of medical education is certainly not easy. In this general discussion section, we have made a critical appraisal of our studies, and we have made many suggestions for improvements and new strategies. Research on medical education has a great future.
References


<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
</tr>
</thead>
</table>
Summary

Patient Feedback in General Practice Training
Chapter 1. General introduction of the thesis

The first chapter describes the context of the research. The importance of adequate consultation skills for practicing general practitioners is undisputed, but coherent guidelines for implementation of teaching these skills among vocational institutes for general practice seem to be lacking. This has consequences for educational developments, innovations and research on education. It is followed by an explanation of the rationale of the research, in which the choice for a patient feedback intervention is motivated. Patient-centeredness and involvement in GP vocational training, and assessment and feedback within a self-directed study environment, are important principles on which this thesis is based. The studies in this project aim at different aspects of evidence for the effectiveness of patient feedback. The chapter ends with an outline of the thesis, and the general aim of the research: to acquire evidence for the effectiveness of patient feedback in the vocational training for general practice.

Chapter 2. Feasibility of the patient feedback programme: a qualitative approach

In order to develop an attractive and effective patient feedback learning programme for general practice trainees (GPTs), we carried out an exploratory study followed by a feasibility study focusing on first-year GPTs and their patients. In the feasibility study among all first-year GPTs at the VU medical center in Amsterdam, the GPTs collected 878 feedback questionnaires from their patients. The extent to which patient feedback was valuated by GPTs, patients and staff, was qualitatively assessed (group discussions and evaluation questionnaires). Although the GPTs and patients alike considered patient feedback to be a major source for the acquisition of consultation skills, the reported learning experiences were more limited than we had expected because of the flattering answers the patients gave and because it were mainly the challenging consultations that provided the potential for relevant patient feedback. Furthermore, there was a preference that the teachers, and not the researchers, should organise and monitor the programme. These findings
formed the basis for recommendations to enhance the learning potential of the patient feedback programme.

Chapter 3. Validity and reliability of the Patient Feedback questionnaire on Consultation skills (PFC)

This study focused on the validation of a patient feedback questionnaire, the PFC, the aim of which was to assess (patient-centered) consultation skills in the subsequent phases of a consultation. An existing questionnaire concerning patient-centeredness (PPPC) was adapted to cover the ‘communicator’ items in the CANMEDS competency profile. The PFC adequately covers the corresponding ‘communicator’ competency (face and content validity), assessed by interviewing patients and GPTs during the developmental process. The content validity of the PFC was determined by experts (n=10). Based on the data from 222 PFCs, factor analysis showed a one-dimensional construct. The internal consistency was high (Cronbach’s alpha 0.89). For the single items, the response rate varied from 89.2% to 100%; and the maximum score (ceiling effect) varied from 45.5% to 89.2%. The PFC appears to be a valid, single-scale, internally consistent instrument, that may be a valuable learning tool with which GPTs, other physicians and medical students can acquire feedback from patients with regard to their consultation skills. It has the potential to contribute to the acquisition of communication skills by GPTs, it is applicable in most consultations, has acceptable ceiling effects, and it takes only a few minutes to complete.

Chapter 4: A comparison of patients’ and doctors’ views on consultation skills

In a cross-sectional study based on 888 pairs of patient-doctor combinations, we investigated the agreement between patient and doctor perspectives on components of patient-centeredness: ‘exploring the disease and illness experience’, ‘finding common ground’, and ‘personal context’. Therefore, 48 first-year GPTs and their patients completed the Patient Perception of Patient-Centeredness (PPPC) questionnaire (item-score range: 1-4). The mean item-scores were 3.63 (SD 0.38) and 2.94 (SD 0.49), for patient and GPT
assessments, respectively (p<0.0001). The relatively low scores for ‘finding common ground’ (3.54 [SD 0.54]) and ‘personal context’ (3.17 [SD 0.96]), as compared to ‘exploring the disease and illness experience’ (3.78 [SD 0.34]), indicate that these components should receive more attention in the vocational training of GPTs. Multilevel analysis showed that younger patients, patients of male gender, greater consultation complexity, and Dutch as a ‘second’ language of GPTs, were related to lower patient assessments. Studying the differences in these aspects of consultation skills might form a basis for new learning strategies for GPTs.

Chapter 5. A controlled trial to assess the effectiveness of patient feedback on consultation skills

The aim of this controlled trial was to assess whether an intervention, i.e. an additional patient feedback training programme, leads to better consultation skills of GPTs than regular communication training, and whether process measures (intensity of participation in the programme) predict the effect of the intervention. In 2006, first-year GPTs at the VU medical center were allocated into either an intervention group (n=23) or a control group (n=30). Standardised simulation patients visited the practices and video-taped their consultations at baseline and after three months. The video-tapes were randomly assigned to eight trained staff-members, and assessed by means of the MAAS-Global instrument (range 0-6). Data on 50 GPTs were available for the follow-up analysis (intention-to-treat analysis). GPTs in both the intervention group and the control group improved their consultation skills: the mean MAAS-Global baseline scores were 3.29 (SD 0.75), and the mean follow-up scores were 3.54 (SD 0.66) for all participants (p=.047). The improvement in MAAS-Global scores in the intervention group did not differ significantly from the improvement in the control group. Multilevel, linear mixed model analysis showed a trend that the intensity of participation in the patient feedback programme predicted more improvement in the MAAS-Global scores. Although the baseline scores were already in the high range, consultation skills in both groups showed a moderate improvement. This is reassuring for current teaching methods.
Chapter 6. Systematic review of the learning potential of patient feedback

In the systematic review described in this chapter we analysed the evidence for the educational potential of real-patient feedback on consultation skills of practicing physicians. Therefore, five electronic databases were searched. All empirical studies of all study designs (randomized controlled trials, quasi-experimental, cross-sectional and qualitative designs) were eligible for inclusion. Articles were assessed for level of educational impact (from physicians’ valuation assessment of the learning experience to assessment of actual change). Eighty percent of the 15 studies that were included showed an improved outcome (effect) as a result of patient feedback. However, positive results were more common among studies assessing a lower level of educational impact than among studies assessing actual change in consultation skills (considered the highest level of educational impact). In conclusion, there is evidence for the educational potential of patient feedback on physicians’ consultation skills, however, the evidence for actual improvement is limited.

Chapter 7. Factors influencing the participation of general practice trainees in a patient feedback programme

This study builds further on the evidence from the trial, i.e. that GPTs who actively participated improved their consultation skills more than those who did not. It therefore seems relevant to analyse the participation behaviour of the GPTs, which was defined as the number of patients included for feedback, and showed great diversity. We studied possible determinants (among which baseline consultation skills) of participation in a self-directed learning programme with real patient involvement. Furthermore, semi-structured interviews were held with those GPTs who collected less than five questionnaires. Lower baseline consultation skills of GPTs were not related to a higher rate of participation. Among male GPTs and GPTs with more clinical experience, there was a lower participation rate. The GPTs did not mention vulnerability or awkwardness with regard to patient involvement as a reason for low participation, but they had various motivational and practical reasons for questioning the expected benefits of the programme. It is probable that close
monitoring and coaching of individual GPTs by tutors and research staff are important requirements for self-directed learning from patient feedback. This could enhance the motivation of participants, which might ultimately benefit the imbedding of a patient-centered approach in primary care.

Chapter 8. Generalisability of MAAS-Global assessments of consultation skills
The last study described in this thesis focused on aspects of generalisability when teachers, after attending a short training course, assessed the consultation skills of GPTs with the MAAS-Global Instrument. All 176 randomly allocated video-taped consultations (102 with standardised patients and 74 with real patients) were observed by two teachers, independently. The inter-observer variation was the most important component of variance; the difference in MAAS-Global score between the individual assessors was significant. Reliability was estimated with the generalisability coefficient (GC). For reliable assessments (GC>0.7), three assessors and 30 standardised-patient consultations were needed, or two assessors and nine-real-patient consultations. The teachers completed a questionnaire, in which they indicated that they felt sufficiently competent in using the MAAS-Global instrument. Therefore, for reliable assessment of the consultation skills of GPTs, multiple observations are required, as indicated by the results of this study. We recommend the assessment of real-patient consultations instead of consultations with standardised patients.

Chapter 9. General discussion
This chapter contains a synthesis of the findings of the studies described in this thesis, focusing on implications for the vocational training for general practice and recommendations for further research. It explains that there are several levels of educational impact of patient feedback, which has consequences for future decisions to implement patient feedback in training programmes. Furthermore, it urges the need for a close collaboration between researchers, teachers and practicing GPs to generate and conduct educational research.
Samenvatting

Patiënten feedback in de opleiding huisartsgeneeskunde
Hoofdstuk 1. Algemene introductie van het proefschrift
Het eerste hoofdstuk betreft de context van het onderzoek. Hoewel het belang van adequate consultvaardigheden van praktiserende huisartsen algemeen is aanvaard, ontbreken eenduidige en coherente richtlijnen en onderwijsmethoden tussen de verschillende opleidingsinstituten huisartsgeneeskunde. Dit heeft ook gevolgen voor ontwikkelingen en innovaties in het vlak van communicatie-onderwijs en onderwijsgebonden onderzoek. De keuze voor patiënten feedback op de consultvaardigheden van huisartsen in opleiding als rationale van het onderzoek wordt uitgelegd. Patiënt-gerichtheid, patiënten participatie in de huisartsopleiding, beoordeling en feedback, zijn uitgangswaarden waarop dit proefschrift is gebaseerd. De diverse studies in dit onderzoeksproject richten zich op diverse onderdelen van de educatieve waarde van patiënten feedback, en vormen de bouwstenen van het bewijs er van. Het hoofdstuk eindigt met een uiteenzetting van de inhoud van het proefschrift en een beschrijving van het algemene onderzoeksdoel: het beoordelen van de effectiviteit van patiënten feedback in de huisartsopleiding.

Hoofdstuk 2. Het patiënten feedback onderwijsprogramma: een kwalitatieve benadering
Voor de ontwikkeling van een attractief en effectief patiënten feedback programma voor huisartsen in opleiding hebben we een verkennende studie en een haalbaarheidsstudie uitgevoerd bij eerstejaars huisartsen in opleiding en hun patiënten. In de haalbaarheidsstudie verzamelden alle eerstejaars huisartsen in opleiding van het VU medisch centrum te Amsterdam 878 feedback vragenlijsten van hun patiënten. Het kwalitatieve aspect van dit onderzoek (groepsdiscussies en evaluatie van de vragenlijsten) stelde de mate vast waarin patiënten feedback werd gewaardeerd door huisartsen in opleiding, de patiënten en de staf van de huisartsopleiding. Hoewel zowel de huisartsen in opleiding als de patiënten de feedback als een belangrijke bron beschouwden voor het aanleren van consultvaardigheden, werden de gerapporteerde leermogelijkheden beperkt door de geflatteerde antwoorden die patiënten gaven, en door de beperking dat slechts de wat ingewikkelder consulten
voldoende potentieel boden voor relevante patiënten feedback. Er werd een voorkeur uitgesproken dat docenten, en niet de onderzoekers, het onderwijs gaven en organiseerden, omdat zij onderwijsdoelen prefereerden boven onderzoeksdoelen. Deze bevindingen vormden de basis voor aanbevelingen om het leerpotentieel van het patiënten feedback programma te vergroten.

**Hoofdstuk 3: Validiteit en betrouwbaarheid van de Patient Feedback questionnaire on Consultation skills (PFC)**

De studie beschreven in dit hoofdstuk is gericht op de validering van de patiënten feedback vragenlijst, de PFC. Deze vragenlijst heeft als doel om de aspecten van patiëntgerichtheid van consultvaardigheden te meten in de opeenvolgende stadia van het consult. Een bestaande vragenlijst betreffende patiënt-gerichtheid (de PPPC) diende als basis en werd aangepast om meer onderdelen van het CANMEDS competentieprofiel ‘communicatie’ te kunnen omvatten.

De PFC komt inhoudelijk overeen met de ‘communicatie’ competentie (inhouds-validiteit), hetgeen bleek uit interviews met patiënten en huisartsen in opleiding gedurende de ontwikkelingsfase. De inhouds-validiteit van de PFC werd bevestigd door experts (n=10). Op basis van de data van 222 PFCs, werd middels factor analyse een één-dimensionaal construct aangetoond. De interne consistentie van de PFC was hoog (Cronbach’s alpha 0.89). Het responspercentage varieerde van 89.2 tot 100 voor wat betreft de afzonderlijke vragen; het percentage patiënten dat de maximum score invulde (plafond effect) varieerde voor de afzonderlijke vragen van 45.5 tot 89.2. De PFC bleek een valide, enkelvoudig geschaald, intern consistent instrument te zijn, dat waardevol kan zijn voor huisartsen in opleiding voor het aanleren van consultvaardigheden. Het is toepasbaar bij de meeste consulten, heeft acceptabele plafond effecten, en het invullen er van kost slechts enkele minuten.
Hoofdstuk 4. Patiënten- en artsen perspectief op consultvaardigheden
In een dwarsdoorsnede-onderzoek gebaseerd op 888 arts-patiënt combinaties, werd het patiënten perspectief onderzocht met betrekking tot componenten van patiënt-gerichtheid: het exploreren van de ziekte en de ziektebeleving; het overeenstemming vinden met de patiënt, en het informeren naar de persoonlijke context. De patiënten van 48 eerstejaars huisartsen in opleiding vulden de Patient Perception of Patient-Centeredness (PPPC) questionnaire (item-score range: 1-4) in. De gemiddelde item-score was 3.63 (SD 0.38). De huisartsen in opleiding beoordeelden ook de patiënt-gerichtheid van hun eigen consulten (met een gemiddelde score van 2.94 [SD 0.49], p<0.0001). De relatief lage scores voor ‘overeenstemming vinden’ (3.54 [SD 0.54]) en ‘persoonlijke context’ (3.17 [SD 0.96]), ten opzichte van het ‘exploreren van de ziekte en de ziektebeleving’ (3.78 [SD 0.34]), zijn indicaties dat deze onderdelen meer aandacht zouden behoeven in de huisartsopleiding. Multilevel analyse liet zien dat 1) de jongere groep patiënten, 2) mannelijke patiënten, 3) een grotere complexiteit van het consult, en 4) Nederlands als tweede taal van de huisarts in opleiding gerelateerd zijn aan lagere patiënt beoordelingen. Het bestuderen van de verschillen tussen arts- en patiënt beoordelingen betreffende de consulten kan een basis vormen voor nieuwe leerstrategieën voor huisartsen in opleiding.

Hoofdstuk 5. Een gecontroleerde trial voor het beoordelen van de effectiviteit van patiënten feedback op de consultvaardigheden
Het doel van deze gecontroleerde trial was het vaststellen of een interventie, te weten een toegevoegd patiënten feedback programma, tot betere consultvaardigheden van huisartsen in opleiding zou leiden in vergelijking met de bestaande communicatievaardigheden training. De patiënten feedback vragenlijst (PFC) werd hierbij gebruikt als onderwijs-instrument (niet als beoordelingsinstrument). Tevens werd onderzocht of procesmaten, in het bijzonder de intensiteit van deelname aan het programma, het effect van de interventie beïnvloeden.

In 2006 werden alle eerstejaars huisartsen in opleiding van het VU medisch centrum toegewezen aan ofwel een interventie groep (n=23), ofwel een
controle groep (n=30). Gestandaardiseerde simulatie-patiënten bezochten de huisartspraktijken en maakten video-opnames van de consulten als uitgangswaarde en na drie maanden. De video-opnames werden gerandomiseerd verdeeld over acht getrainde stafleden (docenten), en beoordeeld met behulp van de MAAS-Globaal (range 0-6). Vijftig huisartsen in opleiding waren beschikbaar voor vervolganalyse (intention-to-treat analyse). Huisartsen in zowel de interventie als de controle groep verbeterden hun consultvaardigheden: de gemiddelde MAAS-Globaal uitgangsscore was 3.29 (SD 0.75), en de gemiddelde vervolg score was 3.54 (SD 0.66), voor alle deelnemers (p=.047). De verbetering in MAAS-Globaal scores in de interventie groep verschilde niet significant van de controle groep. Multilevel, lineaire mixed model analyse liet een trend zien dat de intensiteit van de deelname aan het patiënt feedback programma een grotere verbetering voorspelde in de MAAS-Globaal scores. Hoewel de uitgangsscores zich al in de hoge range bevonden, toonden beide studiegroepen een aanzienlijke verbetering, hetgeen bemoedigend is voor de huidige onderwijsmethoden.

Hoofdstuk 6. Systematische review betreffende het leerpotentieel van patiënten feedback

De systematische review beschreven in dit hoofdstuk betreft de analyse van het bestaande bewijs voor de educatieve waarde van feedback van echte patiënten op de consultvaardigheden van praktiserende artsen. Vijf elektronische databanken werden doorzocht. Alle empirische studies van diverse soorten onderzoeksdesign (gerandomiseerde gecontroleerde trials, quasi-experimentele, dwarsdoorsnede- en kwalitatieve designs) kwamen in aanmerking voor inclusie. Artikelen werden beoordeeld op niveau van educatieve impact (van waardering voor het programma, tot toegenomen kennis, tot aangetoonde verandering in gedrag). Van de 15 geïncludeerde artikelen vond 80% een verbetering in de uitkomst (effect) ten gevolge van patiënten feedback. Deze positieve resultaten werden echter voornamelijk gevonden bij studies met uitkomstmaten op een lager niveau van educatieve impact, in tegenstelling tot studies die zich richten op het niveau van werkelijke gedragsveranderingen met betrekking tot de
consultvaardigheden (beschouwd als het hoogste niveau van educatieve impact). Concluderend: er is bewijs voor de educatieve waarde van patiënten feedback op de consultvaardigheden van artsen, maar het effect op werkelijke verbetering is beperkt.

Hoofdstuk 7. Factoren gerelateerd aan de actieve deelname van huisartsen in opleiding aan het patiënten feedback programma

In deze studie bouwen we verder op de resultaten van de trial, in het bijzonder het gegeven dat huisartsen in opleiding die actief deelnamen aan het onderwijsprogramma hun consultvaardigheden meer verbeterden ten opzichte van anderen die minder of niet participeerden. Bovendien toonde de participatie van huisartsen in opleiding een grote diversiteit. Het is daarom relevant om het deelnamegedrag van de huisartsen in opleiding te analyseren, hetgeen werd gedefinieerd als het aantal patiënten dat zij includeerden voor het verkrijgen van feedback.

Een laag uitgangsniveau van consultvaardigheden was niet gerelateerd aan een grotere deelname aan het patiënten feedback programma. Blijkbaar herkenden de huisartsen hun (relatieve) deficiënties niet, of zetten deze niet om in een actievere deelname. Onder mannelijke en meer ervaren huisartsen in opleiding werd een lagere graad van deelname gevonden.

Semi-gestructureerde interviews werden gehouden met huisartsen in opleiding die minder dan vijf patiënten hadden weten te incluideren. Deze huisartsen in opleiding voelden zich niet kwetsbaar of ongemakkelijk met het betrekken van echte patiënten in de huisartsopleiding. Dit werd ook niet aangevoerd als reden voor een lagere deelname, maar wel werden diverse praktische redenen genoemd en twijfelden enkelen over de te verwachten effectiviteit van het programma. Het is goed mogelijk dat een persoonlijker begeleiding van de (individuele) huisartsen in opleiding door de docenten of onderzoekers gewenst of zelfs een voorwaarde is voor het slagen van zelfsturend leren van patiënten feedback. Dit zal mogelijk de motivatie van de deelnemers verhogen, en uiteindelijk de inpassing van een patiënt-gerichte benadering in de huisartsopleiding ten goede komen.
Hoofdstuk 8. Generaliseerbaarheid van de MAAS-Globaal beoordelingen van consultvaardigheden

De laatste studie beschreven in dit proefschrift is gericht op aspecten van de betrouwbaarheid (generaliseerbaarheid) als docenten, na het volgen van een korte training, de consultvaardigheden beoordelen van huisartsen in opleiding met behulp van de MAAS-Globaal. Alle 176 op video vastgelegde consulten (102 met gestandaardiseerde simulatiepatiënten en 74 met echte patiënten) werden gerandomiseerd verspreid en beoordeeld door twee docenten, onafhankelijk van elkaar. De inter-beoordelaars variatie was de grootste variantiecomponent; het verschil in MAAS-Globaal beoordelingen tussen de docenten onderling was significant. De betrouwbaarheid werd geschat met de generaliseerbaarheid coëfficiënt (GC). Voor betrouwbare beoordelingen (GC>0.7), zijn drie beoordelaars en 30 gestandaardiseerde consulten nodig, of twee beoordelaars en negen consulten met echte patiënten. De docenten vulden een vragenlijst in, waarin werd aangegeven dat ze zich voldoende vaardig achten in het gebruik van de MAAS-Globaal.

Voor betrouwbare beoordelingen van de consultvaardigheden van huisartsen in opleiding zijn verscheidene, zo niet vele observaties noodzakelijk, zoals is aangegeven door de resultaten van deze studie. Voor betrouwbaardere beoordelingen adviseren we het gebruik van consulten met echte patiënten in plaats van gestandaardiseerde.

Hoofdstuk 9. Algemene discussie

Dit hoofdstuk bevat een synthese van de bevindingen van de studies uit dit proefschrift, gefocust op toepasbaarheid in de huisartsopleiding, en bevat aanbevelingen voor toekomstig onderwijsgebonden onderzoek. Het maakt duidelijk dat er verscheidene niveaus zijn waarop het effect van patiënten feedback kan worden vastgesteld, hetgeen van invloed is op toekomstige uitvoeringsplannen. Verder bevat het een pleidooi voor een coherente samenwerking tussen onderzoek, onderwijs en de (academische) huisartspraktijk voor het genereren en uitvoeren van onderzoeksplannen binnen de huisartsopleiding.
Samenvatting
Terugkijkend op de afgelopen jaren, waarin ik me intensief heb bezig gehouden met dit onderzoek, had ik de constructieve samenwerking met alle betrokkenen niet willen missen. Het werken aan dit proefschrift heb ik beschouwd als een voorrecht, en nu wordt het de hoogste tijd al diegenen te bedanken zonder wie het nooit tot stand zou zijn gekomen. Ik heb veel van jullie kunnen leren.

Al voor mijn aanstelling als onderzoeker werden plannen gesmeed voor een patiënten feedback onderzoek op de huisartsopleiding VUmc. Nettie Blankenstein en Harm van Marwijk, co-promotoren, jullie zijn tijdens alle jaren van uitstekende samenwerking het dichtst betrokken gebleven bij de voortgang en wisten wat een promovendus nodig heeft. Onder het motto ‘Hak je eigen hout, dan word je twee keer warm’ hebben jullie me niets cadeau gedaan, maar waren wel altijd bereikbaar. Een warm dankwoord voor jullie beiden!

Wim Stalman, gewezen promotor, nu decaan van de VUmc, je hebt mij binnen gehaald bij het EMGO en enthousiast gemaakt voor dit onderwerp, dat nog niet binnen de bestaande onderzoekslijnen paste. Jouw visie op de academisering van de eerstelijnszorg deel ik van harte; jouw menselijkheid is mij een voorbeeld.

Henriëtte van der Horst, dank voor je promotor-schap, dat je halverwege hebt overgenomen. Je inbreng getuigde van groot wetenschappelijk inzicht. Je opmerkingen op de artikelvoorstellen gaven de aanzet de resultaten nog kritischer te beschouwen, hetgeen de kwaliteit immer ten goede kwam.

Piet Schoonheim, jij hebt aangegeven welke nieuwe ontwikkelingen je belangrijk acht voor de huisartsopleiding en leidt de huisartsopleiding met moderne visie. Je gaf mij de mogelijkheid al tijdens mijn huisartsopleiding een start te maken met het onderzoek, waarvoor dank.

De beslissing om het nieuwe project eerst te laten landen in de huisartsopleiding, en het project te verbeteren aan de hand van de opgedane ervaringen, is een goede geweest.

Hartelijk dank gaat uit naar de beloftevolle en enthousiaste collega’s (en hun opleiders) die hun medewerking aan de pilot verleenden: Daniel Blankenberg, Marjolein Koenders, Els Licht, Anton van de Vusse, Mieke Balk, Jet Heinsbroek,
Chantal van het Hooft, Dana Acherman, Anouk Vrasdonk, Willem Koch, Annerose Kunenborg, Daniel Barten, Diana Seegers, Reinier Groeneveld, Sabine Visser, en natuurlijk alle patiënten, die zonder uitzondering bereid waren mee te werken. Barbara Emmerich, AIOS en stagiaire, heeft aan de ontwikkeling van het project bijgedragen. François Schellevis, Judy van Es, Margit Vermeulen, Barend van Duin, Bas Maiburg, Ferry Bastiaans en Giel Nijpels gaven hun oordeel over de vorm en inhoud van de vragenlijst.


Het slagen van onderwijsgebonden onderzoek berust grotendeels op een degelijke methodologie. Riekie de Vet, hoogleraar klinimetrie, je hebt een steeds grotere en belangrijkere rol ingenomen als adviseur in het project, geweldig. Dirk Knol, aan de meeste hoofdstukken heb je een belangrijke, zo niet onmisbare bijdrage geleverd. Je was bereid vele uren te stoppen in de statistische onderbouwing van de hoofdstukken.

Na een jaar van opgedane ervaring ging het dan echt van start. De eerstejaars AIOS van 2006 waren bereid deel te nemen aan de interventie- of controlegroepen en de bijbehorende metingen met de simulatiepatiënten. Opvallend veel goede consultvoerders gezien! Hartelijk dank, allen, voor jullie betrokkenheid. Dank aan de acteurs die de rol van simulatiepatiënt zo goed vervulden: Roelant Radier, Ellen van Rossum, Amanda Wolzak. Vanuit de
toneelwereld zo maar beland in de medische wereld, jullie beschouwden de opdracht echt als een missie en reisden het hele land door. Piet Schoonheim bepleitte het idee om de docenten van de huisartsopleiding te betrekken in het beoordelen van de onderzoeksbanden, een voorbeeld voor de integratie van onderzoek en onderwijs! De docenten die met veel toewijding alle opnames zorgvuldig beoordeelden waren Anouk Bogers, Willem Feijen, Jolijn de Graaff, Simone van den Hil, Gerrit Locher, Leni Overmars, Harry Schleypen en Gerda Visser. Paul Ram, jij was bereid deze groep te trainen op je bekende enthousiasmerende wijze. Jae Klaasen hielp bij het digitaliseren van de grote verscheidenheid aan bandjes, en Irene Veerhuis maakte de elektronische database beschikbaar. Allen, oprechte dank!

Voor het hoofdstuk over de betrouwbaarheid van de metingen werd samengewerkt met de Universiteit van Maastricht. Cees van der Vleuten en Ron Hoogenboom, jullie bijdragen waren zeer waardevol. De systematische review in dit hoofdstuk is op informele wijze voorzien van advies door Daniëlle van der Windt, nu werkzaam als hoogleraar in Keele, en door Petra Jellema. Ilse Jansma heeft een niet aflatende inspanning geleverd met het maken van een passende zoekstrategie. Ik ben bevoorrecht met jullie bijdrage! Bridget Ryan, thank you for being such a fine co-author. It was a great opportunity to meet you at the NAPCRG congress in Montreal.

Leden van de leescommissie:
Paul Ram, gedurende het onderzoekstraject kwamen onze wegen diverse keren samen en heb ik van je wijsheid kunnen profiteren. Het goede nieuws is dat aan onze samenwerking nog geen einde komt! Moira Stewart: dear Moira, you have played an important role in this thesis. Not only have you developed the concept of patient-centered care, which has been such an important basis for this research, you also were a kind host when you invited us in London, Ontario, in Canada. You have honoured us with your visit to oppose in the public defence of this thesis.
Nawoord

Paul van Royen, uw voordrachten tijdens de EGPRN cursus in Malmö in 2006 over een gedegen methodologische fundering van kwalitatief onderzoek waren inspirerend voor dit onderzoek; dank, ook voor uw opponeren.

Overige leden van de leescommissie: Gerda Croiset, Caroline Terwee en Henk de Vries: hartelijk dank voor jullie commentaar op het manuscript. De beide laatsten, en Fedde Scheele, dank voor jullie bereidheid te willen opponeren.

Wilko Berghoef, mijn huisartsopleider eerste jaar: jij hebt mij op het rechte pad van de huisartsgeneeskunde gebracht. Tijs Stolk, huisartsopleider derde jaar, jij hebt een grote bijdrage geleverd aan mijn professionele ontwikkeling. Ik mag mij gelukkig prijzen met twee zulke goede opleiders!


Faith Maddever, dank voor je correcties op het Engels. Taal behoort tot het fijngevoelige instrumentarium van de wetenschap, en daardoor heb je een belangrijke bijdrage aan de publicaties geleverd.
Nawoord

De Commissie Activiteiten Stimulering Huisartsopleiding (CASH) van de Stichting Beroepsopleiding Huisartsen heeft een prachtige financiële bijdrage geleverd en vertrouwen getoond in de goede afloop, met behoud van volledige wetenschappelijke onafhankelijkheid. Veel dank ben ik verschuldigd en ik hoop dat nog vele huisartsen en/of promovendi zich bij jullie mogen melden.

Beste paranimfen, Martin Landheer en Ron Onstenk, vrienden sinds het begin van de geneeskunde studie. Vroeger was het verstandig je als promovendus te laten beschermen door twee sterke kerels tegen uit de hand lopende academische discussies. Jullie weten wat je te doen staat!

Lieve familie, schoonfamilie en vrienden, jullie zorgden voor een liefdevolle basis en stonden klaar als de wetenschap en de realiteit van alledag (weer) eens niet met elkaar door een deur konden.

Lieve Krista, onze grap was altijd dat zonder jou dit proefschrift veel eerder klaar zou zijn geweest. Je weet dat je dit als een groot compliment mag beschouwen! Lieve Daan, lieve Susanne, ik ben zo trots op jullie!
About the author

Marcel Reinders was born on September 16, 1965 in Amsterdam, the Netherlands. He graduated from secondary school (GSGD, Doetinchem) in 1984. In 1992 he graduated from the School of Medicine, University of Amsterdam. In the first few years, he has worked as a transplant coordinator, and participated in experimental research. Later on, he has worked as a clinical doctor (surgery, urology, geriatrics) in various hospitals and institutes. From 2003 to 2005 he attended the vocational training institute for general practice at the VU medical center. Since then he worked as a general practitioner in Haarlem and area.

During the last stage of his general practice training he started a research pilot on patient feedback. At the end of 2005 the project was officially acknowledged as a PhD project and financially supported by the Dutch Foundation for the Vocational Training of General Practitioners. He completed the post-graduate master programme in epidemiology at the EMGO+ Institute for Health and Care Research in 2010. He recently started his work at the National Vocational Training Institute for General Practice in Utrecht.

He is married to Krista Miedema, who is a cardiologist, and they have a son (Daan, 2000) and a daughter (Susanne, 2002); they live in Haarlem.