

Feedback for General Practitioners in Training: Quality, Styles, and Preferences

FRANS J. PRINS*, DOMINIQUE M.A. SLUIJSMANS and PAUL A. KIRSCHNER

*Open Universiteit Nederland, P.O. Box 2960, 6401 DL, Heerlen, The Netherlands (*author for correspondence, Phone: +31-45-5762292; Fax: +31-45-5762907; E-mail: f.j.prins@fss.uu.nl)*

Received 19 April 2005; accepted 17 September 2005

Abstract. A General Practitioner (GP) is no longer a loner, but a team player in either a group practice or a care centre. This change has led to a concomitant growth in curricular interest in skills essential for successful collaboration and for enhancing critical reflection towards colleagues' performance. Giving and receiving constructive feedback are examples of these skills. The aim of this study was to gain insight in the style and quality of feedback reports on consultation skills written by GPs-in-Training (GPiTs) and by their GP-trainers. Furthermore, the preferences of the GPiTs concerning feedback style were examined. Results show significant differences between GP-trainers and GPiTs in feedback style and quality. A ranking task indicated that GPiTs have a preference for reports characterised by a large number of reflective remarks. Questionnaire results indicate the added value of the use of peer feedback. Implications for the integration of peer feedback activities in the curriculum of GPiTs are discussed.

Key words: consultation skills, feedback preferences, feedback quality, feedback style, peer feedback

A competent General Practitioner (GP) commands a large repertoire of declarative, procedural, and strategic knowledge that they apply in complex cognitive and motor skills (i.e., performance) relating to how they deal with patients (General Medical Council, 2001). To acquire these skills, GPs-in-training (GPiTs) must go through three distinct skill development phases (Anderson, 1982; Fitts, 1964; Lane, 1987; Van Merriënboer, 1997). In the *cognitive/declarative phase*, the learner collects facts, background information, and general rules related to the skill. Performance is slow and effortful, requiring full attention and the learner is dependent on external cues, prompts and formative feedback on the correctness of the performance. The learner leaves this phase with a basic understanding of the task requirements and a set of strategies that is not fully elaborated or integrated. In the *associative/knowledge compilation phase*, the skill is refined. New patterns of skill components are tried out and inappropriate actions are gradually eliminated; the learner begins to 'know' how to act in the required manner. Practice makes

skills more polished and easier to apply. Feedback here not only informs the learner of correctness, but also explains how and why things went wrong or could have gone better. Furthermore, it can allow for discussion, argumentation with others, and internal reflection (Kirschner, 1991). In the third and final phase, the *autonomy/proceduralisation phase*, the learner gains speed, control, and coordination, and eventually achieves skilled performance. Here, feedback is often direct and aimed primarily at correctness. This is usually determined internally by the learner and not via teacher, tutor, or other external person.

Reaching skilled performance (phase three) is for the most part determined by the type and amount of *practice* (Neves and Anderson, 1981; Snoddy, 1926) and *feedback* on that practice (Butler and Winne, 1995; Mory, 1992). While practice receives much attention in medical training, giving and receiving systematic feedback on practice is somewhat underexposed. This article focuses on the role that feedback can play in the education of GPiTs, specifically the style and quality of feedback reports made by GPiTs and GP-trainers, the development of feedback skills of GPiTs, and their feedback preferences. In the presented study, feedback has a formative function; it is directed at improving medical skills.

In medical education, feedback for GPiTs could be provided by GP-trainers as well as fellow GPiTs. In the latter case, one speaks of *peer feedback*. Topping (1998) defines giving and receiving peer feedback as an “arrangement in which individuals consider the amount, level, value, worth, quality or success of the products or outcomes of learning of peers of similar status” (p. 250). During peer feedback, participants are observed by their peers while they are performing. The level of proficiency demonstrated is analysed and reported on and is then compared with predetermined performance standards. All participants have the opportunity to attain the standards, and can thus play a role in giving constructive feedback to their peers.

Organising peer feedback activities is valuable for the professional development of GPs and professionals in general. Sluijsmans et al. (1999) studying applications of peer feedback in several higher education contexts found that peer feedback increases confidence in the ability to perform, awareness of one’s own quality, and reflection in one’s own behaviour and/or performance. They also found that peer feedback improves the quality of learning and independence and responsibility of those who were peer evaluated. These conclusions are in line with Norcini (2003), who provides an interesting overview of peer feedback practices in medical education. He found that although participants had some difficulties in understanding the peer feedback procedures because they were mainly focused on the ranking and scoring of peers, they generally valued its use. Calhoun et al. (1990) note that professional development, growth of student responsibility and autonomy, and development of collaborative attitudes towards faculty staff as the main benefits of peer feedback.

Peer feedback skills have been added in the behaviour repertoire of GPs. The University of Pennsylvania (2004), for example, expects residents to be able to recognise and identify deficiencies in peer performance. Residency programme directors at the University Hospital and Clinics at the University of Wisconsin at Madison identified nine generic abilities that a medical graduate should possess to be admitted into residency training. These abilities included appraisal, analysis, and assessing own and peer performance and inter-personal skills (Stone, 1996). This is not strange if one takes into account (see Table I) that in the last decade the number of GPs working in a group practice in the Netherlands has increased by more than 250% and in duos by 15% while the number of GPs working alone has decreased by 14% (CBS, 2003). GPs in group practices work with each other's patients, fill-in for each other, and often spend considerable time and effort forming a well-functioning team. Thus, peer GPs must be able to constructively criticise each other's behaviour so that the general quality of the group practice is increased and the differences between the doctors and how they deal with patients are minimised.

Many promising results of peer feedback interventions in education have been noted (Sluijsmans, 2002). However, in literature issues concerning the *quality* of feedback, *feedback style*, and *feedback preferences* of both the giver and the receiver are still not systematically elaborated on. It is crucial that feedback, whether provided by a peer or a GP-trainer, meets a number of *quality* criteria to be effective and acceptable. Sluijsmans (2002) advocates that fostering reflection and improvement of performance based on the peer feedback requires certain skills in both the giver and the receiver of the feedback. Most important is that both the giver and the receiver need to understand the appropriate criteria for the performance that is observed. For a GP this means that the criteria for determining what a proper consultation is, for example, must be understood and interpreted correctly. The importance of defining criteria is stressed in many studies (e.g. Falchikov, 1995; Mehrens et al., 1998; Orsmond et al., 1996). Calhoun et al. (1990) concluded that medical students' feedback became progressively more accurate as the course progressed, and that their exposure to clinical work enabled them to better internalise and interpret assessment criteria. Along with having

Table I. Overview of number of GPs working in solo, duo, and group practices

	1995	2000	2001	2002	2003
GPs	4775	4801	4750	4680	4631
Solo practice	3343	3158	3059	2960	2866
Duo practice	1065	1197	1210	1197	1213
Group practice	217	446	481	523	552

appropriate criteria, the feedback giver must possess the knowledge and skills required to adequately present feedback (see Butler and Winne, 1995; Mory, 1992, 2003 for guidelines) and should be able to translate the feedback in a written and/or oral report that is transparent enough for the receiver to improve her/his performance. Recent research has shown that intensive training is required to acquire these skills, and that both educators and educational scientists underestimate the complexity of the process of providing and receiving adequate feedback (Sluijsmans, 2002).

Along with its quality, feedback is also characterised by a certain *style*. Studies in different feedback styles are rare. Van den Berg (2003), following Lockhart and Ng (1995), distinguished four types of feedback: (1) *authoritative*, where the provider gives feedback without explanations or suggestions for revision; (2) *interpretive*, where the provider sticks to her/his own experience and ideas when giving suggestions for improvement; (3) *probing*, where the provider takes the perspective of the receiver and explains her/his remarks, structures feedback according to the performance and provides suggestions for performance improvement; and (4) *collaborative*, which resembles probing, but where the provider and receiver create a collaborative product. This latter position, of course, is not possible in cases where learners work on individual products. One could state that the third and fourth feedback positions are good examples of styles in which a balance between the feedback giver and feedback receiver is established.

Finally, the *preference* of participants in feedback style is a topic of research interest. Participants involved in peer feedback activities have been shown to have both positive and negative experiences with peer feedback, depending on the feedback procedure followed (Cheng and Warren, 1997; Conway et al., 1993; Williams, 1992). It is important to study whether preference for a specific feedback style is beneficial for an effective peer feedback process (Prins et al., 2005). The match between provider style and receiver preference could be important determinants for the effectiveness of the feedback.

This study attempts to identify the style and quality of current feedback practices of post-graduate GPiTs and their GP-trainers, and the feedback preferences of the feedback receivers (i.e., GPiTs). The physician/patient encounter was chosen as the object of feedback, since such encounters are at the centre of a GP's functioning. The GP needs to act in many different situations with many different people with many different problems and becoming proficient at conducting such encounters requires considerable practice.

As stated earlier, guidance in efficient and effective peer feedback skills is not explicitly included in most medical curricula, despite the acknowledged need for it (e.g., Rudy et al., 2001). The Faculty of Medicine at the University of Maastricht (The Netherlands) expressed a desire to include

peer feedback activities in its GP-curriculum with the goal of (1) achieving a higher involvement of GPiTs in the discussion about the criteria for constructive feedback to support competency-development in carrying out physician/patient encounters, and (2) getting more insight in the personal perceptions of GPs about several aspects of their professional development. To address these goals, a study was set up with the following specific research questions:

1. What is the style and quality of the feedback reports written by GP-trainers and GPiTs?
2. What are the preferences of GPiTs concerning feedback style?
3. What are the perceptions of GPiTs concerning their role as GPs, the value of peer feedback, and their self-efficacy in giving and receiving feedback?

Methods

PARTICIPANTS

Participants were 46 GPiTs (25 first-year, 21 third-year post-master level) and 12 GP-trainers. Note that all GPiTs in the Netherlands have completed a 5-year master programme in general medicine and work 4 days a week in a group practice under supervision of an experienced GP. One day per week is focused on educational activities and group meetings where daily problems are discussed and videotapes are sometimes presented. These group meetings are led by GP-trainers.

DESIGN AND PROCEDURE

Three data sets were collected. First, in separate sessions, six groups of 8–12 GPiTs ($n = 46$) and 12 GP-trainers were asked to write a qualitative feedback report for a video recording of a physician/patient encounter. The GPiTs had to imagine that the GP in the video was one of their peers and the GP-trainers had to imagine that the GP in the video was a GPiT in the group that they supervise.

Second, four discriminating feedback reports were selected from the sessions with the GP-trainers. The GPiTs were asked to use a ranking form to rank these feedback reports based upon their own personal preferences and to explain their preferences. This ranking task was followed by an open discussion to identify differences between the GPiTs about presenting and receiving feedback and the ethical issues involved in presenting and receiving feedback.

Finally, the GPiTs filled out a questionnaire concerning their role as GP, and feedback.

INSTRUMENTS

Videotaped consultation

Cases involving videotaped physician/patient encounters can be valuable for both the acquisition of consultation skills and for the use of peer feedback (Lane and Gottlieb, 2004; Ram et al., 1999). To identify feedback strategies of GP-trainers and GPiTs, a 6-minute video registration of a simulated physician/patient encounter was used. In this video, a GP could be observed examining a patient with a sore throat.

Rating forms for style and quality of feedback reports

Instruments developed in prior studies (e.g., Sluijsmans et al., 2002) were used for developing instruments for assessing style and quality of peer feedback reports. Three variables determined the style and quality, namely use of criteria, nature of the feedback, and writing style.

The first rating form categorised the statements made by the feedback provider in her/his report, and was aimed at the assessment of the *feedback style* (quality was not taken into account here). *Use of criteria* was categorised by the number of: criteria used, medical remarks, remarks concerning physician/patient communication, descriptions of behaviour, and explanations of remarks. *Nature of the feedback* was categorised by the number of: positive remarks, negative remarks, reflective questions, external examples, and suggestions for improvement. *Writing style* was categorised by: structure, use of key words or descriptions, and use of first person.

To analyse the *quality* of the feedback reports, a scoring rubric was used in which nine items could be scored on the 'Feedback Quality Index'. Two items were related to the use of criteria (i.e., quality of both content and explanations; 50% of the score), four were related to the nature of the feedback (i.e., remarks, questions posed, repertoire, and advice; 35% of the score), and three were related to the quality of writing (i.e., structure, formulation, and style; 15% of the score). A score could be given for each item (see Table II). The scoring range for feedback quality was between 0 and 100.

Two judges (the first two authors) independently scored one third of the feedback reports on style using the rating form, and on quality using the rubric. For each variable, Cronbach's α was calculated over the scores of the judges as an inter-judge reliability coefficient. These reliabilities were acceptable for all variables, with an average of 0.84 and a minimum of 0.74. The scores of the first judge were used for the analyses.

Table II. Scoring rubric for the quality of feedback reports

Main category	Sub category	Good achievement	Average achievement	Minimal achievement
1. Criteria	• Content	Substantial medical and doctor-patient communication related remarks	30 Some medical and some doctor-patient communication related remarks	15 No or hardly any medical and doctor-patient communication related remarks
	• Explanations	Description of behaviour and explanation of remarks throughout the report	20 Some descriptions of behaviour and some explanations of remarks	10 No description of behaviour and no explanation of remarks
	• Remarks	Balanced number of positive and negative remarks	10 Positive remarks dominate	5 Negative remarks dominate
2. Nature	• Posed questions	Questions fostering reflection throughout the report	10 Some questions that stimulate reflection	5 No questions in the report
	• Repertoire	Good external examples (e.g., own experiences)	5 Unclear examples	2 No examples
3. Writing style	• Advice	Good and clear suggestions for improvement; constructive advice	10 Some suggestions for improvement	5 No suggestions for improvement; no constructive advice
	• Structure	Clear structure (e.g., chronology)	5 Unclear structure	2 No structure
	• Formulation	Short descriptions	5 Key words dominate	2 Only key words
	• Style	First person throughout the report	5 Sometimes first person	2 No first person, judging

Ranking form for preference of feedback reports

To determine the GPiTs' preferences concerning feedback reports, four discriminating feedback reports were selected from sessions with the GP-trainers. Report A was structured by *positive and negative* remarks, key words dominated, used first person, and asked reflective questions. Report B was not structured, had a *descriptive* style, used first person, and contained many reflective questions, examples and suggestions for improvement. Report C was the shortest of the four, lacked a consistent structure, had a descriptive style, did not use first person, had no reflective questions and contained only one suggestion for improvement. Finally, report D had a chronological structure, a descriptive style, used first person, had neither reflective questions nor examples, and contained some suggestions for improvement. The GPiTs used a ranking form to rank the four feedback reports according to their preferences on a scale of 1 (most preferred) to 4 (least preferred) and to explain their ranking.

GPiT questionnaire

This questionnaire had broader educational purposes and concerned GPiT's perceptions on their role of GP, their vision on instruction and assessment, feedback, and ethical issues. Ninety-two items dealt with 12 variables, but only 7 variables had a relation with feedback and will be described here. The variables were self-efficacy concerning their role as GP (9 items, $\alpha=0.73$, example "As a GP, I can deal efficiently with unexpected events"), self-efficacy concerning giving medical consultations (7 items, $\alpha=0.63$, example "I am good at carrying out a good medical consultation"), self-efficacy concerning feedback (10 items, $\alpha=0.71$, "I can give understandable feedback"), perceived added-value of giving feedback (4 items, $\alpha=0.77$, example "Giving each other feedback is instructive"), perceived added-value of receiving feedback (6 items, $\alpha=0.66$, example "I can learn from receiving feedback"), feedback affect – feeling safe and confident when giving and receiving feedback – (13 items, $\alpha=0.88$, example "I experience receiving feedback as threatening"), and ethical aspects relating to the fairness and transparency of feedback (8 items, $\alpha=0.64$, example "Generally, the feedback that I receive is fair"). Perceived self-efficacy is defined (Bandura, 1994) as people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives and determines how people feel, think, motivate themselves, and behave. The items of the questionnaire were 5-point Likert-scale, varying from "I totally disagree" to "I totally agree".

Results

STYLE AND QUALITY OF THE FEEDBACK REPORTS

A principal component analysis was performed on measures of feedback style of GPiTs and GP-trainers ($n = 58$) to explore dimensions of feedback style. Four components with an eigenvalue greater than 1.0 were extracted. The first component had an eigenvalue of 3.22, which accounted for 29.3% of the variance. The eigenvalue of the second component was 2.05 (18.7% of the variance), the third was 1.17 (10.7%), and the fourth was 1.04 (9.4%). Table III presents the unrotated component matrix. The first component can be interpreted as *use of content-based criteria*, and the second as *nature of feedback and writing style*. Components 3 and 4 are very similar (loads for negative remarks and reflection), but differ with respect to explanations (positive load on component 3 and negative on 4) and medical remarks (loads only on component 4). Component 3 could be considered a *formative* dimension (i.e., telling what is wrong, offering explanations, and stimulating reflection) while component 4 is more *summative* (i.e., telling what is wrong, making medical remarks, stimulating reflection, but not giving any explanations).

No significant differences between first and third-year GPiTs were found for feedback style and quality variables, thus, means and standard deviations are reported for the total GPiT sample. Table IV presents the means and standard deviations concerning feedback style for GP-trainers and GPiTs. Differences between GP-trainers and GPiTs were significant for several variables concerning feedback style. GP-trainers used more criteria in their feedback report ($t(56) = 3.12, p < 0.01$), made more remarks concerning GP-patient communication ($t(56) = 4.36, p < 0.01$), described actual behaviour

Table III. Unrotated component matrix for measures of feedback style (loadings > 0.40)

	Component 1	Component 2	Component 3	Component 4
Criteria used	0.82			
Medical remarks				0.42
Communication remarks	0.91			
Descriptions of behaviour	0.90			
Explanations	0.43		0.53	-0.52
Positive remarks	0.73			
Negative remarks		-0.57	0.41	0.48
Reflective questions			0.63	0.44
Examples mentioned		0.65		
Suggestions for improvement		0.64		
Use of first person		0.75		

Table IV. Means and standard deviations concerning feedback style for GP-trainers and GPiTs

Variables	GP-trainers ($n = 12$)		GPiTs ($n = 46$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Criteria used*	10.25	3.62	7.30	2.71
Medical remarks	0.83	0.94	0.67	0.90
Communication remarks*	9.42	6.44	4.70	1.93
Descriptions of behaviour*	7.50	6.45	2.72	1.86
Explanations	1.33	0.98	0.70	1.17
Positive remarks*	5.00	2.73	3.28	2.07
Negative remarks	1.17	2.29	1.43	1.78
Reflective questions	1.50	1.62	0.83	1.20
Examples mentioned	0.50	1.24	0.11	0.31
Suggestions for improvement	1.08	0.79	0.96	1.11
Use of first person*	3.00	2.80	0.61	1.02

* $p < 0.01$.

more often ($t(56) = 4.46$, $p < 0.01$), made more positive remarks ($t(56) = 2.39$, $p < 0.01$), and used first-person style more often ($t(56) = 4.79$, $p < 0.01$) than GPiTs. They did not use significantly more remarks to stimulate reflection. Furthermore, the GP-trainers' feedback reports were generally of a significantly better quality than those made by the GPiTs ($M_{\text{GP-trainers}} = 71.83$, $SD_{\text{GP-trainers}} = 15.09$; $M_{\text{GPiTs}} = 45.63$, $SD_{\text{GPiTs}} = 15.63$; $t(56) = 5.21$, $p < 0.01$).

Frequency analyses revealed that 5 out of 12 GP-trainers (42%) did not ask reflective questions, 3 (25%) gave no suggestions for improvement, and 7 feedback reports of the GP-trainers (58%) were unstructured. There were individual differences in which particular techniques the GP-trainers used, that is, no GP-trainer failed to use all of them. For the GPiTs, 28 out of 46 GPiTs (61%) did not support their remarks with explanations, 22 (48%) did not use negative remarks, 27 (59%) did not ask a single reflective question, 20 (44%) gave not a single suggestion for improvement, 30 (65%) did not use first person in their report, and 39 reports (85%) had no structure. Apparently, not all GP-trainers and GPiTs have these skills available or do not spontaneously use them while writing a feedback report.

PREFERENCES OF GPiTs CONCERNING FEEDBACK STYLE

The quality score and the mean rank number for report A was 82 and 2.61 ($SD = 1.11$), respectively, for report B 90 and 1.65 ($SD = 0.92$), for report C 65 and 3.57 ($SD = 0.72$), and for report D 80 and 2.17 ($SD = 0.71$). A one-way ANOVA, with report as independent variable (46 observations for each

report) and rank number as dependent variable, showed that differences between reports were significant ($F(3,180)=39.06, p<0.01$). A *post-hoc* Bonferroni test revealed that differences between all four reports were significant except for the difference between report A and D. These results show that GPiTs have a sense of feedback quality and prefer feedback reports that are descriptive and contain many reflective remarks and a personal style (report B).

GPiT QUESTIONNAIRE

No significant differences between first and third-year GPiTs were found for the questionnaire variables, thus the means and standard deviations of the 7 variables are reported for the total GPiT sample. The results indicated that the GPiTs perceive themselves as capable of coping with situations that a GP encounters ($M=3.53, SD=0.44$) and coping with medical consults ($M=3.48, SD=0.48$). GPiTs also clearly stated that giving ($M=4.20, SD=0.52$) and receiving feedback ($M=4.28, SD=0.34$) has added value. Moreover, they themselves feel capable of giving feedback ($M=3.81, SD=0.35$) and they feel moderately safe in a feedback situation ($M=2.46, SD=0.63$). Finally, GPiTs consider the feedback they receive as fair and clear ($M=3.68, SD=0.37$).

Some of the variables of the GPiT questionnaire were related. First, the added value of giving feedback and the added value of receiving feedback were positively related ($r=0.61, p<0.01$). Moreover, variables concerning self-efficacy were positively related. Self-efficacy with respect to medical consults correlated 0.70 with self-efficacy for their role as GP ($p<0.01$) and 0.45 with self-efficacy for giving and receiving feedback ($p<0.01$). In addition, feedback affect correlated negatively with the three measures of self-efficacy: -0.54 ($p<0.01$) with self-efficacy GP, -0.67 ($p<0.01$) with self-efficacy medical consultation, and -0.53 ($p<0.01$) with self-efficacy feedback. GPiTs with low self-efficacy do not feel safe in feedback situations.

Some measures of style, quality, and preferences were related to questionnaire variables. The number of reflective questions in the feedback report correlated negatively with self-efficacy with their role as GP ($r=-0.37, p<0.05$), negatively with self-efficacy towards medical consultations ($r=-0.51, p<0.01$), negatively with self-efficacy for giving and receiving feedback ($r=-0.34, p<0.05$), and positively with feedback affect ($r=0.30, p<0.05$). It appears that the less capable or unsafe one feels in feedback situations, the more reflective questions will be used in a feedback report. For measures of preferences, rank number of report C correlated negatively with self-efficacy with respect to the role as GP ($r=-0.33, p<0.05$), and with self-efficacy with respect to medical consultations ($r=-0.41, p<0.01$).

Perhaps GPiTs who feel that they can deal effectively in the job-situation prefer short feedback reports without reflective remarks.

Conclusions and Discussion

The aims of this study were to examine style and quality of feedback reports of GP-trainers and GPiTs, and to examine GPiTs' feedback preferences and perceptions. Results showed that GP-trainers outperformed GPiTs in feedback style and quality. Moreover, although GPiTs clearly stated that giving and receiving feedback has added value, that they feel capable of doing so, and that the feedback they receive is fair and clear, the style and quality of their feedback reports definitively warrants improvement. In general, GPiTs used a rather small set of criteria in their reports and the majority of both groups delivered feedback reports without structure and with limited stimulation for reflection; the reports contained hardly any reflective questions, suggestions for performance improvement, or examples. GPiTs preferred feedback characterised by a descriptive style, first person use, many reflective questions, examples and suggestions for improvement. The preferred feedback report served both functions of feedback: correcting errors and fostering reflection in the receiver's mind.

The low quality of the GPiTs' peer-feedback reports is reason for concern. GPiTs may not demonstrate adequate feedback behaviour because they fail to spontaneously enact the appropriate available strategy (i.e., *production deficiency*), or because they do not have the appropriate skill available in their repertoire (i.e., *availability deficiency*) (Flavell, 1976; Veenman et al., 2000). To increase the quality of the peer feedback, instructional support or training aimed at giving feedback is warranted. Instructional support can alleviate *production deficiency* by prompting specific feedback activities available to the GPiT. Feedback instruments such as performance scoring rubrics with criteria, or structured feedback forms that force feedback providers to ask reflective questions and give suggestions for improvement could be valuable instruments for increasing the quality of the peer feedback. However, when GPiTs demonstrate an *availability deficiency*, instructional support will not be sufficient and a systematic training for acquiring feedback skills will be more appropriate. Training such complex skills takes a substantial amount of deliberate practice and thus should be integrated into the curriculum (Sluijsmans, 2002).

In this study, no differences were found on either feedback style or quality between first-year and third-year GPiTs, which is a strong indication of inadequate development of feedback skills during the academic career of the GPiTs. To remedy this, training is needed for acquiring feedback skills which focuses on the use of criteria, the nature of feedback, and writing style, and which consists of specific feedback assignments supported by feedback

instruments. After peer feedback training, GPiTs should be able to spontaneously provide good quality peer feedback, even without supporting instruments. In other words, feedback training should solve both production and availability deficiencies.

The skill 'giving good feedback' is an important asset for a competent GP-trainer, as is observation and analysis skills, and the ability to foster reflection in trainees (Boendermakera et al., 2000). The results reported in the research presented here show that the style and quality of the feedback of GP-trainers should also be improved. They need to better stimulate reflection by asking more reflective questions, providing more examples, and providing more suggestions for performance improvement. An effective approach for increasing and guaranteeing the quality of the GP-trainers' feedback is the use of collegial consultancy – exposing the GP-trainers to feedback by their own colleagues – complemented by a training on (peer) feedback. Based on the results of this study, participating GP-trainers showed a willingness to change and discuss issues concerning feedback.

Results of the principal component analysis on the variables of feedback style suggest that feedback style has more than one dimension, and that the first two dimensions more or less resemble the feedback skills identified by Sluijsmans (2002). The first dimension may be interpreted as *use of criteria*, which is content related, while the second may be interpreted as related to *nature of the feedback and writing style*. The third dimension could be considered a *formative* one (i.e., telling what is wrong, offering explanations, and stimulating reflection) while the fourth dimension is more *summative* (i.e., telling what is wrong, making medical remarks, stimulating reflection, but not giving any explanations). Thus, feedback support and training should be aimed at three of the four dimensions, not necessarily with same instruments or tasks and at the same time. If the assessment is formative, that means it should aim at dimensions 1, 2, and 3; if it is summative, then it should aim at dimensions 1, 2, and 4.

Just as 'beauty is in the eye of the beholder', the value of feedback might lie in the 'perception of the receiver'. Prins et al. (2005), agreeing with Mory (2003) suggest research to "identify measurable variables that can reflect internal cognitive and affective processes of learners that might potentially affect how feedback is perceived and utilised" (p. 777). The present study showed that GPiTs – as feedback receivers – have a sense of feedback quality and that individual differences in preferences were small with the majority preferring feedback that was descriptive, reflective, and personal. The next step is to help feedback receivers take an active role in the feedback process, for example by equipping them to ask for particular types of feedback, to determine whether the feedback is clear, whether they agree, and to request explanations and suggestions for improvement. This way, feedback is coconstructed through loops of dialogue and information (Askew and

Lodge, 2000) and the collaborative feedback position (Lockhart and Ng, 1995; Van den Berg, 2003) is within reach. Feedback receivers should be convinced that they themselves are responsible for getting the feedback they need to improve performance.

Receiving feedback that meets one's preferences, however, does not guarantee its effectiveness. What we like is not always the same as what we need. Future research should examine the relation between preferences and effectiveness of feedback in an experimental design. Effectiveness may concern the improvement of performance and the improvement of feedback skills.

The results of this study provide important guidelines for effectively implementing peer feedback in the medical curriculum. By taking the preferences of the feedback receiver as well as the development of the receiver's active role into account GPiTs can develop their feedback skills and gradually become valuable participants in the professional development process of their colleagues.

References

- Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review* **89**: 369–406.
- Askew, S. & Lodge, C. (2000). Gifts, ping-pong and loops – Linking feedback and learning. In: S. Askew (ed.), *Feedback for Learning*, London: Routledge/Falmer, pp. 1–17.
- Bandura, A. (1994). Self-efficacy. In: V.S. Ramachaudran (ed.), *Encyclopedia of Human Behavior*, Vol. 4, pp. 71–81. New York: Academic Press (Reprinted in H. Friedman (ed.), *Encyclopedia of Mental Health*. San Diego: Academic Press, 1998).
- Boendermaker, P.M., Schulinga, J., Meyboom-de Jong, B., Zwierstrac, R.P. & Metz, J.C.M. (2000). What are the characteristics of the competent general practitioner trainer? *Family Practice* **7**(6): 547–553.
- Butler, D.L. & Winne, P.H. (1995). Feedback and self regulated learning: A theoretical synthesis. *Review of Educational Research* **65**: 245–281.
- Calhoun, J.C., Ten Haken, J.D. & Woolliscroft, J.O. (1990). Medical students' development of self- and peer assessment skills: A longitudinal study. *Teaching and Learning in Medicine* **2**: 25–29.
- CBS (2003). *Vademecum gezondheidsstatistiek Nederland 2003* [Vademecum health statistics 2003]. Voorburg, The Netherlands: Centraal Bureau voor de Statistiek, Ministerie van Volksgezondheid, Welzijn en Sport.
- Cheng, W. & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education* **22**: 233–239.
- Conway, R., Kember, D., Sivan, A. & Wu, M. (1993). Peer assessment of an individual's contribution to a group project. *Assessment and Evaluation in Higher Education* **18**: 45–56.
- Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International* **32**: 175–187.
- Fitts, P.M. (1964). Perceptual-motor skill learning. In: A.W. Melton (ed.), *Categories of Human Learning*. London: Academic Press.
- Flavell, J.H. (1976). Metacognitive aspects of problem solving. In: L.B. Resnick (ed.), *The Nature of Intelligence*. NJ: Erlbaum, Hillsdale, pp. 231–236.
- General Medical Council (2001). *Good Medical Practice*. London: General Medical Council.
- Kirschner, P.A. (1991). *Practicals in Higher Science Education*. Utrecht: Lemma.
- Lane, N.E. (1987). *Skill Acquisition Rates and Patterns: Issues and Training Implications*. New York: Springer.

- Lane, J.L. & Gottlieb, R.P. (2004). Improving the interviewing and self-assessment skills of medical students: Is it time to readopt videotaping as an educational tool? *Ambulatory Pediatrics* **4**: 244–248.
- Lockhart, C. & Ng, P. (1995). Analyzing talk in ESL peer response groups: Stances, functions, and content. *Language Learning* **45**(4): 605–655.
- Mehrens, W.A., Popham, W.J. & Ryan, J.M. (1998). How to prepare students for performance assessments. *Educational Measurement: Issues and Practice* **17**(1): 18–22.
- Mory, E.H. (1992). The use of informational feedback in instruction: Implications for future research. *Educational Technology Research and Development* **40**: 5–20.
- Mory, E.H. (2003). Feedback research revisited. In: D.H. Jonassen (ed.), *Handbook of Research for Educational Communications and Technology*. New York: Macmillan Library Reference, pp. 745–783.
- Neves, D.M. & Anderson, J.R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In: J.R. Anderson (ed.), *Cognitive Skills and Their Acquisition*. Hillsdale, NJ: Erlbaum, pp. 57–84.
- Norcini, J. (2003). Peer assessment of competence. *Medical Education* **37**: 539–543.
- Orsmond, P., Merry, S. & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment and Evaluation in Higher Education* **21**: 239–249.
- Prins, F., Sluijsmans, D.M.A., Kirschner, P. & Strijbos, J.W. (2005). Formative peer assessment in a CSCL environment: A case study. *Assessment and Evaluation in Higher Education* **30**: 417–444.
- Ram, P., Grol, R., Rethans, J.J., Schouten, B., van der Vleuten, C. & Kester, A. (1999). Assessment of general practitioners by video observation of communicative and medical performance in daily practice: Issues of validity, reliability and feasibility. *Medical Education* **33**: 447–454.
- Rudy, D.W., Fejfar, M.C., Griffith, C.H. III & Wilson, J.F. (2001). Self- and peer assessment in a first-year communication and interviewing course. *Evaluation & the Health Professions* **24**: 436–445.
- Sluijsmans, D.M.A. (2002). *Student Involvement in Assessment. The Training of Peer Assessment Skills*. Unpublished doctoral dissertation, Open University of the Netherlands, Heerlen.
- Sluijsmans, D.M.A., Brand-Gruwel & Van Merriënboer, J.J.G. (2002). Peer assessment training in teacher education. *Assessment and Evaluation in Higher Education* **27**(5): 443–454.
- Sluijsmans, D.M.A., Dochy, F. & Moerkerke, G. (1999). Creating a learning environment by using self-peer- and co-assessment. *Learning Environments Research* **1**: 293–319.
- Snoddy, G.S. (1926). Learning and stability. *Journal of Applied Psychology* **10**: 1–36.
- Stone, H.L. (June 1996). *Lessons Learned from FIPSE Projects III*. University of Wisconsin at Madison. An Ability-Based Assessment program at the Medical School. Project director: Howard L. Stone, Medical Sciences Center, University of Wisconsin-Madison, 1300 University Avenue, Madison, WI 53706. <http://www.ed.gov/offices/OPE/FIPSE/LessonsIII/madis.html>.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research* **68**(3): 249–276.
- University of Pennsylvania (2004). Department of Medicine Internal Medicine Residency Curriculum Academic Year 2004–2005.
- Van den Berg, I. (2003). *Peer assessment in universitair onderwijs* [Peer assessment in university education]. Unpublished doctoral dissertation, University of Utrecht, The Netherlands.
- Van Merriënboer, J.J.G. (1997). *Training Complex Cognitive Skills: A Four-Component Instructional Design Model for Technical Training*. Englewood Cliffs, NJ: Educational Technology Publications.
- Veenman, M.V.J., Kerseboom, L. & Imthorn, C. (2000). Test anxiety and metacognitive skillfulness: Availability versus production deficiencies. *Anxiety, Stress, and Coping* **13**: 391–412.
- Williams E. (1992). Student attitudes towards approaches to learning and assessment. *Assessment and Evaluation in Higher Education* **17**: 45–58.