

Structured feedback to undergraduate medical students: 3 years' experience of an assessment tool

Ann-Christin Haffling, Anders Beckman & Gudrun Edgren

To cite this article: Ann-Christin Haffling, Anders Beckman & Gudrun Edgren (2011) Structured feedback to undergraduate medical students: 3 years' experience of an assessment tool, *Medical Teacher*, 33:7, e349-e357, DOI: [10.3109/0142159X.2011.577466](https://doi.org/10.3109/0142159X.2011.577466)

To link to this article: <https://doi.org/10.3109/0142159X.2011.577466>



Published online: 22 Jun 2011.



Submit your article to this journal [↗](#)



Article views: 2740



View related articles [↗](#)



Citing articles: 3 View citing articles [↗](#)

WEB PAPER

Structured feedback to undergraduate medical students: 3 years' experience of an assessment tool

ANN-CHRISTIN HAFFLING, ANDERS BECKMAN & GUDRUN EDGREN

Lund University, Sweden

Abstract

Background: There is a paucity of research on the effects of interactive feedback methods and sustained assessment strategies in formative assessment of students in the workplace.

Aims: To investigate the outcome of long-term use of an assessment tool.

Methods: Retrospective analysis of summarised assessment tools from 3 years of 464 final-year students in general practice. Quantitative data were analysed using non-parametric tests and a multi-level approach. Qualitative data were subjected to content analysis.

Results: Students' main deficiencies in the consultation were in the domains of working diagnoses and management plans; however, supervisors emphasised goals of patient-centred communication and structure of the medical interview. As a group, students underestimated their clinical performance, compared to supervisors' judgement. Most students were supplied with specific goals, 58% with specific follow-up feedback. The majority of students and supervisors were satisfied with the assessment strategy. Long-term experience with the tool significantly increased the proportion of specific goals and feedback to students, supervisors' stringency of the assessment, and their satisfaction with the tool.

Conclusions: The summarised assessment strategy proved feasible and acceptable with students and supervisors in a continuous attachment with assigned personal supervisors. However, there was room for improvement in supervisors' provision of specific follow-up feedback.

Introduction

Feedback on students' performance in clinical settings is a fundamental component in students' development of clinical skills (Ende 1983; Cantillon & Sargeant 2008). Feedback should focus on action, not on the individual; be specific, and linked to personal goals, to help reinforce desirable performance and to correct poor performance (Hewson & Little 1998; Hattie & Timperley 2007; Cantillon & Sargeant 2008; Archer 2010). The qualities of effective feedback are supported by research (Hewson & Little 1998; Hattie & Timperley 2007), and are associated with students' perceptions of high-quality teaching (Cantillon & Sargeant 2008).

Work-based assessment with feedback can have a powerful impact on the change of medical students' behaviour (Norcini & Burch 2007). However, several studies suggest that direct observation and formative assessment of students in the workplace occur rather infrequently (Daelmans et al. 2004; Howley & Wilson 2004; Kogan & Hauer 2006; Pulito et al. 2006). Furthermore, the feedback provided is often of questionable value to students (Daelmans et al. 2006; Kogan & Shea 2008).

Poor supervisor contribution has been identified as the most important limiting factor in work-based assessment and feedback (Norcini & Burch 2007). By involving supervisors in

Practice points

- In a continuous attachment with personal supervisors, summarised formative assessment is a feasible and acceptable strategy.
- Most students and supervisors are satisfied with the experience of a structured assessment tool.
- Long-term use of an assessment strategy can increase the proportion of specific feedback to students, supervisors' stringency of assessment, and their satisfaction with the tool.

the planning of the process, the frequency of assessment and feedback can increase (Daelmans et al. 2006), and by offering training to supervisors in rating students' performance, accuracy can improve (Shumway & Harden 2003). To facilitate assessment of students and residents in the workplace, several unique structured tools have been developed and implemented (Kogan et al. 2009). Most tools include items on communication, history taking, physical examination, management and counselling.

The reliability of work-based assessment is a matter of appropriate sampling across content, patients and assessors, but in formative assessment a compromise on reliability in

Correspondence: A.-C. Haffling, Clinical Research Centre (CRC), Entrance 72, Malmö University Hospital, 205 02 Malmö, Sweden. Tel: +46 40 391374; fax: +46 40 391370; email: ann-christin.haffling@med.lu.se

favour of educational impact is acceptable (van der Vleuten 1996). However, the educational impact of assessment and feedback on students' performance is an issue that has only been sparsely researched (Norcini & Burch 2007). Students' positive perceptions of the assessment process are reported in several studies (Lane & Gottlieb 2000; McKinley et al. 2000; Paukert et al. 2002; Burch et al. 2006), but results from students' modifications of knowledge, attitudes and skills, self-assessed or assessed by their supervisors, are more seldom described (Kogan et al. 2009).

Interactive feedback methods, including students' self-assessment and an action plan for improvement with follow-up, are critical components of effective feedback (Hattie & Timperley 2007), but underutilised in feedback to students (Fernando et al. 2008) and to residents (Holmboe et al. 2004). Written educational strategies for improvement (Hastings et al. 2006), and students' self-assessment with feedback comments from supervisors and patients (Braend et al. 2010) have been reported. However, these studies did not include follow-up on feedback provided.

Long-term use of an assessment tool might allow supervisors to gradually get accustomed to the tool and improve in their assessment and feedback to students. However, most studies on workplace-based assessment of undergraduates are based on short interventions, and there is a paucity of research reports on sustainable assessment with feedback as a routine feature in the workplace (Norcini & Burch 2007).

The medical school curriculum of Lund University, Sweden, comprises five and a half years (11 semesters). At the time of the study, the intake was approximately 80 students twice a year. The programme of Community Medicine ran in the second half of students' 10th semester, and included 16 days (4 days every second week) of practice in a health centre. Students' end-of-clerkship evaluations and results from an investigation of the educational climate, measured by DREEM (Edgren et al. 2010), raised concerns about the feedback process during the general practice attachment. To improve the provision of feedback on students' performance, we introduced an assessment and feedback tool, which included students' self-assessment, mutually agreed goals, and follow-up feedback.

This article retrospectively analyses the results from the first 3 years' use of the tool. The principal aim of this study was to investigate the outcome of long-term use of an assessment tool; specifically the content and nature of feedback provided, supervisors' evaluations of students' clinical performance and achievements, students' self-assessment, and students' and supervisors' attitudes to the feedback strategy.

Methods

Setting

The objectives of the general practice attachment are that students will learn to manage common diseases and complaints in primary care, supervised mainly by personal general practice supervisors, but also by other general practitioners (GPs) at the practice. GPs in 50–60 health centres are engaged as supervisors. Students are expected to independently

Box 1. Competency domains of the assessment form.

Medical interview
History taking
Physical examination
Working diagnoses
Problem solving
Investigations and treatment
Explanation and planning
Relationship with patients; time management

perform medical interviews and physical examinations, present working diagnoses and management plans, and conclude consultations under supervision. Supervisors are requested to frequently observe students during medical interviews and physical examinations, and are supposed to provide brief feedback daily as part of the flow of work. Moreover, summarised, formal feedback should be given twice during the attachment, at the midpoint and at the end.

In Sweden, medical students are allowed to work as locums in hospital after nine completed semesters in medical school. They work as 'interns', however closely supervised and with limited duties and responsibilities. Several students make use of this opportunity, particularly during summer holidays.

Assessment tool

We designed a tool for mandatory formative assessment and feedback to be applied in the two summarised feedback sessions. Prior to the study we had used evaluation forms, focusing mainly on students' communication skills and with poor opportunities for assessment of other domains of the consultation, i.e. clinical reasoning, problem solving and management. The new tool was inspired by the modified version of the Leicester Assessment Package (LAP), developed for general practice, and previously found to be valid and reliable (McKinley et al. 2000). The modified LAP included five competency domains; however, we altered the instrument slightly, subdividing the three 'double' domains, resulting in eight domains of 'equal value' (Box 1). A seven-point scale was used with a rating span from 1–2 (unsatisfactory), 3 (borderline), 4–5 (satisfactory) to 6–7 (superior). The detailed competencies included in the LAP were not evaluated, but applied on the form to clarify the contents of each domain (Appendix). We briefed the GP supervisors on the tool in a regular supervisors' meeting, and involved them in the modification of the form before implementation.

Identical assessment forms were used for students' self-assessment and for supervisors' formative assessment. Feedback sheets, attached to the forms, supplied detailed information on the assessment process and on how to differentiate between the scores of the rating form. The sheets also had space for narrative comments on agreed goals and feedback, and contained a few questions to students and to supervisors (Box 2).

Assessment process

At the start of the programme, students were informed verbally of the assessment process. Although the assessment was

Box 2. Questions to students and supervisors on the feedback sheets.*To students*

- Did your supervisor correctly identify your strengths and weaknesses in your consultations? (yes, partially, no)
- Did you get specific advice on how to improve your consultations? (yes, partially, no)
- Did your supervisor supply effective feedback in the second feedback session? (yes, partially, no)

To supervisors

- Did the assessment tool contribute to improved structure in the assessment of your student's performance? (yes, partially, no)

formative in nature, students were required to self-assess and to be evaluated by their supervisors. Students were supplied with assessment forms and feedback sheets, all required to be completed and returned to the faculty office by the end of the course.

Before the first feedback session, supervisors and students should independently complete the forms by allocating suitable scores in each of the eight domains. The feedback should be provided as a 'reflective feedback conversation' (Cantillon & Sargeant 2008), goals and strategies for further training should be mutually agreed, and recorded on students' and supervisors' feedback sheets. At the end of the attachment students were supposed to be provided with follow-up feedback, documented on both feedback sheets. However, recording on only one of the sheets could be accepted.

As the assessment was formative, the scores of the assessment forms did not contribute to final examination results, but provided a framework for a dialogue between students and supervisors. The final summative examination consisted of a global, non-graded, pass/fail assessment of students' performance and a case-based written test at the end of the semester.

There was no requirement of prior training for supervisors; however, we realised a need for training in two areas: the rating of students' performance, and the provision of effective feedback. We offered a 1-day workshop on effective feedback on four different occasions during the first 2 years of the study. The workshops were based on video recorded role-plays, using written scenarios and standardised students, reviews of videotaped performances, and facilitated small group discussions. Written information with an overview of the basic tenets of effective feedback was also sent to supervisors each semester. Moreover, we included 2 h of rating practice, based on the tool, in our regular supervisors' meetings twice a year, attended by approximately 70% of supervisors.

Ethics permission was at the time of the study not required for research based on data already collected for the purpose of assessment, if data were fully anonymised.

Data analysis

Assessment forms and feedback sheets from students and supervisors were collected during the first 3 years. Data from the first feedback session consisted of scores on the assessment forms and narrative agreed goals on the feedback sheets; from the second session, the narrative follow-up feedback on the feedback sheets.

All quantitative data, including students' results on the final written test, were entered into SPSS, version 17.0. Students and

supervisors were allocated code numbers to prevent identification of individuals. Non-parametric tests (Mann-Whitney U and Kruskal Wallis) were used for the comparison of groups. However, in calculations where groups of students, based on supervisors' scores, were compared, we had to correct for the dependencies with repeated measures, as several students were 'clustered' into the same supervisors. We then used a multi-level approach, where a correction for dependence is built into the model (Maas & Snijders 2003). MIWin version 2.17 (Rasbash et al. 2009) was used for analysis, with residual (or restricted) maximum likelihood.

We estimated the internal consistency reliability of the tool using Cronbach's alpha for each semester and for the whole period of study. Spearman's correlation coefficient was calculated for associations between supervisors' assessment of students' average scores and students' results on the written test. Since the maximum scores of the written test varied between semesters, students' scores were ranked into deciles each semester. Supervisors' assessment of students' average scores were ranked into deciles for the whole period of study.

The narrative comments on the feedback sheets were written verbatim into a word processor by one of the authors, and then subjected to 'basic' content analysis (Crabtree & Miller 1999), involving a categorisation of words and meanings, using predefined templates. Mutually agreed goals in the first feedback session and follow-up feedback in the second session were identified as generalised or specific. Specific goals and specific feedback comments were mapped against the framework of the eight competency domains. The specific feedback was further evaluated for the possible occurrence of 'feed forward', advice for the future (Hattie & Timperley 2007).

Results

Students and supervisors

Analyses were based on the 464 students, who finished the course during the first six semesters (3 years). The number of students each semester varied between 71 and 83; 256 (55%) of students were female; median age was 26 years (range 24–45 years). Half of the students, with equal gender distribution, had worked as locums, and 281 students (61%) had a female supervisor.

The total number of supervisors engaged was 151, varying between 65 and 78 supervisors each semester; 87 (58%) were female. A few supervisors were responsible for more than one student each semester, but in alternate weeks. In the six semesters 49 supervisors were responsible for 4–12 students, 55 for 2–3 students and 47 for only 1 student. A total of 42

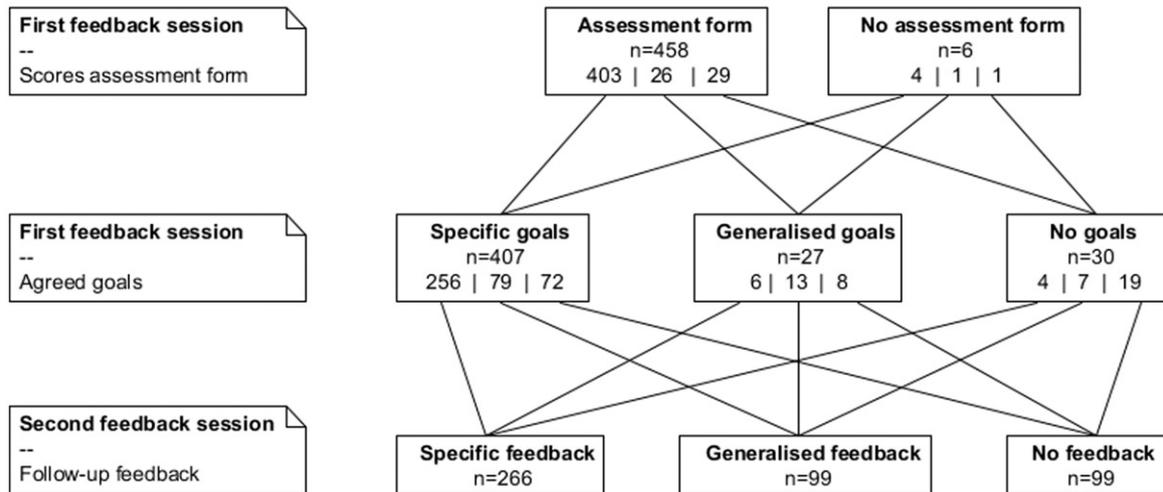


Figure 1. Flow chart of students between first and second feedback sessions.

Table 1. First feedback session: scores of the assessment form.

Competency domain	Mean of ratings					
	Supervisors' ratings (all students) (n = 458)	Self-ratings (all students) (n = 462)	Self-ratings (male students) (n = 206)	Self-ratings (female students) (n = 256)	Self-ratings (students with work experience) (n = 226)	Self-ratings (students without work experience) (n = 228)
Medical interview	5.4 (SD = 1.0)***	5.0 (SD = 0.8)	4.9 (SD = 0.8)	5.0 (SD = 0.8)	5.0 (SD = 0.8)	4.9 (SD = 0.8)
History taking	5.3 (SD = 1.0)***	4.6 (SD = 0.9)	4.6 (SD = 0.9)	4.6 (SD = 0.9)	4.7 (SD = 0.8)	4.5 (SD = 1.0)
Physical examination	5.3 (SD = 1.0)***	4.5 (SD = 1.0)	4.6 (SD = 1.0)	4.5 (SD = 1.0)	4.6 (SD = 1.0)*	4.4 (SD = 1.0)
Working diagnoses	5.0 (SD = 1.0)***	4.2 (SD = 1.0)	4.3 (SD = 1.0)*	4.1 (SD = 1.0)	4.3 (SD = 0.9)	4.1 (SD = 1.0)
Problem solving	5.2 (SD = 1.1)***	4.5 (SD = 1.0)	4.5 (SD = 1.0)*	4.4 (SD = 0.9)	4.5 (SD = 1.0)*	4.4 (SD = 1.0)
Investigations and treatment	5.1 (SD = 1.1)***	4.1 (SD = 1.0)	4.2 (SD = 1.0)*	4.0 (SD = 1.0)	4.2 (SD = 1.0)**	4.0 (SD = 1.0)
Explanation and planning	5.3 (SD = 1.0)***	4.5 (SD = 1.1)	4.6 (SD = 1.0)	4.5 (SD = 1.1)	4.6 (SD = 1.1)	4.5 (SD = 1.1)
Relationship with patients; time management	5.4 (SD = 1.1)***	4.5 (SD = 1.0)	4.5 (SD = 1.0)	4.5 (SD = 1.1)	4.5 (SD = 1.1)	4.5 (SD = 1.0)
Average score	5.3 (SD = 0.9)***	4.5 (SD = 0.8)	4.5 (SD = 0.7)	4.5 (SD = 0.8)	4.5 (SD = 0.7)*	4.4 (SD = 0.8)

Notes: * $p < 0.05$, ** $p < 0.01$; *** $p < 0.001$.

supervisors (28%) took part in one of the four workshops provided.

First feedback session: scores of the assessment form

The assessment form for judging students' performance was used for 458 students; 99% (Figure 1). Six students from the first two semesters were not evaluated, but were provided with free text summaries by their supervisors. Self-assessment data were missing from two students, and data on previous work as locums for eight students.

'Relationship with patients/time management' and 'Medical interview' were awarded supervisors' highest ratings of students' performance, and 'Working diagnoses' and 'Investigations and treatment' the lowest (Table 1). Students' highest self-ratings were assigned to 'Medical interview' and 'History taking', and the lowest to 'Investigations and treatment' and 'Working diagnoses'.

An 'average' score, averaging each student's results in the eight competency domains, was calculated (Table 1). Supervisors' mean of students' average scores was 5.3 (range 3–7), and students' self-assessments showed a mean of the average scores of 4.5 (range 2.4–6.8).

Male students rated themselves significantly higher than female students in three competency domains (Table 1); however, in supervisors' assessment there was no significant difference between genders.

Students with work experience as locums rated themselves significantly higher than students without work experience in three competency domains and in their average scores (Table 1). Supervisors' ratings revealed significantly higher scores for the locum students, compared to the non-locum students in 'Physical examination', 5.5 vs. 5.2 ($p = 0.003$); 'Investigations and treatment', 5.1 vs. 5.0 ($p = 0.02$); and 'Explanation and planning', 5.4 vs. 5.2 ($p = 0.02$).

Most students were evaluated by their supervisors using a range of only two scores, but students themselves spread their scores in a wider range (Table 2).

Table 2. First feedback session: range of scores of the assessment form.

No. of scores	Supervisors (%) (<i>n</i> = 458)	Students (%) (<i>n</i> = 462)
1 score	61 (13)	18 (4)
2 scores	258 (56)	159 (34)
3 scores	123 (27)	218 (47)
4–5 scores	16 (4)	67 (13)

First feedback session: agreed goals

Narrative comments on students' and supervisors' feedback sheets were combined and categorised as specific goals (407 students; 88%; Box 3), generalised goals (27 students; 6%; Box 3) and no goals (30 students, 6%); (Figure 1).

Specific goals were mapped against the framework of the eight competency domains (Table 3). There were 1084 goals recorded in total, with a mean of 2.7 and a range of 1–7 domains per student. The majority of specific goals related to 'Medical interview', particularly three detailed competencies: 'Summarises', 'Negotiates agenda' and 'Elicits patient's perspective'. The domain of 'Explanation and planning' was also frequently addressed, with the emphasis on 'Relates explanations to patient's perspective' and 'Checks patient's understanding'.

Students, supervised by GPs who used a wider range of scores (3–5 scores, compared to 2 scores or 1 score), were significantly more likely to receive specific goals ($p=0.002$). When students' and supervisors' gender was taken into account, female students with female supervisors were provided with specific goals significantly more often (95%); $p=0.004$, compared to female students with male supervisors (85%), male students with female supervisors (85%) and male students with male supervisors (82%).

Second feedback session: follow-up feedback

The narrative comments of the follow-up feedback at the end of practice were categorised into specific feedback (266 students; 58%; Box 3), generalised feedback (99 students; 21%; Box 3) and no feedback (99 students; 21%); (Figure 1). The category of specific feedback was further evaluated, and 86 of these 266 students (32%) were additionally provided with 'feed forward' (Box 3). Of the 266 students, 10 were given specific feedback despite no recorded specific goals (Figure 1).

Specific feedback for the 256 students with previously documented specific goals was mapped against the framework of the eight competency domains (Table 3). There were 729 feedback comments in all, with a mean of 2.8 and a range of 1–6 domains per student. There was no significant difference between genders.

Students' evaluation of the assessment strategy

The response rate for the students' three questions on the quality of feedback varied between 89% and 93%. A total of 77% of respondents were of the opinion that their supervisors had correctly identified their strengths and weaknesses, 20% partially. Concerning specific advice on how to improve

their consultations, 70% of respondents were completely satisfied, 26% partially. Finally, 77% of respondents thought they received effective follow-up feedback, 17% partially.

Students who were provided with specific goals were significantly more satisfied with their supervisors' identification of their strengths and weaknesses (79%), compared to those with generalised goals (72%) and no goals (57%); $p=0.042$. Students who received specific follow-up feedback were significantly more often completely satisfied (81%), compared to those with generalised feedback (78%) and no feedback (63%); $p=0.004$.

Supervisors' evaluation of the assessment strategy

Of the 151 supervisors, 147 (97%) answered the question whether the tool contributed to improved structure in the assessment of students' performance. Several supervisors changed their views from partially to completely positive over the years. Finally, 111 (76% of respondents) were completely positive, 34 (23%) partially and 2 (1%) negative.

Of the 49 supervisors who had been responsible for 4–12 students during the six semesters, 86% were finally completely satisfied, compared to the 55 supervisors who had supervised 2–3 students (80%), and the 43 supervisors who only used the tool once (58%); $p=0.006$. Of the 42 participants of the workshops 86% were completely satisfied.

Time and experience

Specific goals documented at the first feedback session increased significantly between the first three and last three semesters from 84% to 91% of students ($p=0.019$). Specific feedback at the second session improved from 50% of students to 65% between the first three and last three semesters ($p=0.006$).

A significantly increased stringency was found in the assessment, depending on supervisors' experience with the tool. Students ($n=115$) who were rated by experienced supervisors, who were using the tool for the 4th–6th semester, were assigned a significantly lower mean of average score (5.0); $p=0.001$, compared to students ($n=187$) who were supervised by GPs using the tool for the 2nd–3rd semester (5.3) and students ($n=156$) whose supervisors were using the tool for the first time (5.4). However, students assigned to more experienced supervisors did not receive a significantly higher proportion of specific goals or specific feedback.

Internal consistency and correlations with written exam

We estimated internal consistency reliability using Cronbach's alpha for each semester separately (0.94–0.96) and for the whole study period (0.95). Correlations among specific competency domains ranged from 0.63 to 0.83.

The number of students who took the written test at the end of the course was 447; the remaining 17 students took the test on a later occasion. A significant but very weak correlation (Spearman $\rho=0.16$; $p=0.001$) was found between supervisors' ranked average scores of students' clinical performance

Box 3. Examples of generalised/specific goals (first feedback session) and generalised/specific feedback and 'feed forward' (second feedback session).

First feedback session – Generalised goals: examples

- More practical training to increase the 'flow' of the consultation (student 113)
- Improved consultation skills and enhanced efficiency (student 193)
- Consult with many patients to gather maximum of experience (student 301)

First feedback session – Specific goals: examples

- Needs to improve in supplying information in a language understandable to the patient, and in suggesting appropriate follow-up (student 9).
- She has to express more verbal empathy, be sensitive to the patient's non-verbal cues, and ask about the patient's concerns and expectations. Improved disposition of time during the consultation is also required (student 88)
- Summarise the patient's history and negotiate agenda with the patient! (student 328)

Second feedback session – Generalised feedback: examples

- Has made progress (student 105)
- She is a very bright and ambitious student (student 395)
- Very active, independent and positive student (student 425)

Second feedback session – Specific feedback: examples

- He has considerably improved in his problem solving, now adequately applying his theoretical knowledge. The structure of the medical interview and his management suggestions have developed (student 173)
- She has focused on the three agreed areas for training: disposition of the available time, medical decision making (investigations and treatment), and examination technique. The results are satisfying (student 243)
- She has improved in her rapport with patients and in her ability to structure the medical interview (student 359)

Second feedback session – Feed forward: examples

- He has improved, but has to consider the structure of the medical interview, and to summarise for the benefit of himself and the patient (student 50)
- She needs to work more on how to elicit the patient's ideas, concerns and expectations (student 62)
- I could confirm his excellent communication skills, and encouraged him to further improve his examination technique (student 292)

Table 3. First feedback session: specific goals; second feedback session: specific feedback.

Competency domains	First feedback session: No. of specific goals (% of students) (n = 407)	Second feedback session: No. of specific feedback comments (% of students) (n = 256)
Medical interview	188 (46)	141 (55)
History taking	83 (20)	46 (18)
Physical examination	116 (29)	61 (24)
Working diagnoses	137 (34)	92 (36)
Problem solving	126 (31)	85 (33)
Investigations and treatment	153 (38)	103 (40)
Explanation and planning	164 (40)	111 (43)
Relationship with patients; time management	117 (29)	90 (35)

and students' ranked written examination scores. However, this correlation was enhanced when only the 10% students ($n=44$) with the lowest assessment scores were analysed (Spearman $\rho = 0.34$; $p = 0.026$).

Discussion

This observational study on 3 years of work-based assessment, using a retrospective design, explored the whole feedback process. Almost all students were assessed and given feedback in the first session, most students were provided with specific goals, and a majority received follow-up feedback, aligned to their goals. Most students and supervisors were satisfied with the assessment strategy. Long-term use of the tool significantly increased the proportion of specific goals and specific

feedback, supervisors' stringency of the assessment, and their satisfaction with the tool.

Some interesting inconsistencies were found at the first feedback session between scores of the assessment forms and goals agreed for further training. The competency domains awarded the lowest scores by both students and supervisors were 'Working diagnoses' and 'Investigations and treatment', findings consistent with previous research on senior students' performance in general practice (Hastings et al. 2006; Braend et al. 2010). Students probably perceived the generation and testing of hypotheses across the range of non-specific problems in general practice as problematic, and their previous experience from hospital clerkships in suggesting management plans was likely to be sparse. However, these domains, although regarded as students' main deficiencies, only resulted in goals in a third of students. On the other hand, almost half of students were provided with goals in the two domains 'Medical interview' (awarded the highest scores of all domains) and 'Explanation and planning'. Items of patient-centred communication and the structure of the medical interview were particularly emphasised. These goals focused on 'process' before 'content', and were probably perceived as more appropriate by supervisors. Others have also recommended a focus on 'process' teaching in direct observation of students' clinical performance in general practice (Russell 2009).

As a group, students underestimated their clinical performance, compared to their supervisors' judgement; a finding supported by previous research in general practice (Braend et al. 2010). A gender difference was noted, with male students rating their performance significantly higher than females in three competency domains. However, supervisors' ratings did not suggest any gender difference. A trend for higher levels of confidence by males in their self-assessment was also found in the systematic review by Colthart et al. (2008).

There was a significant difference between the locum students' and the non-locum students' self-assessment in three competency domains as well as in their average scores, demonstrating the locum students' enhanced self-assurance. The differences between the locum and non-locum students were supported by supervisors' assessment, particularly in physical examination, where the locum students' performance was evaluated as significantly more skilled. Short periods of locum work evidently resulted in students being more accustomed to and more confident in physical examinations of patients.

As established, self-assessment on its own has clear limitations in medical students' evaluations of their performance, and requires applicable external standards for comparison (Eva & Regehr 2005; Colthart et al. 2008). However, self-assessment as an elicitation of students' perceptions of their performance, before receiving feedback, is a recommended strategy (Hewson & Little 1998; Cantillon & Sargeant 2008). The students in our study could use the feedback for reflection and integration with their self-appraisals, constituting an 'informed self-assessment' (Sargeant et al. 2010). In this process, the reflection, particularly if facilitated by the discussion with the supervisor, could play a key role in deciding whether to accept, assimilate and use the feedback or not (Sargeant et al. 2009).

Narrative comments by supervisors suggested that some supervisors were reluctant to provide honest corrective feedback to students they had worked closely with, which resulted in inflated scores or scores identical for all competency domains. Problems in transmitting unpleasant messages, fear of upsetting the student, or adversely affect their relationship with the student might contribute to supervisors' reluctance. Students with supervisors who used a wide range of scores received significantly more specific goals, a finding supported by previous research (Fernando et al. 2008). Narrative comments by these supervisors indicated that they probably took the time to assess students accurately in the different domains and to discuss specific goals. Students also valued specific goals, designed to improve their performance, and these students were significantly more satisfied with their supervisors' identification of their strengths and weaknesses. Moreover, female students with female supervisors were significantly more likely to receive specific goals, compared to remaining student-supervisor dyads. However, these findings could only permit incomplete conclusions. A previous study showed that female supervisors were more likely to give feedback to male than to female students, and that the gender dyad resulting in most feedback was male supervisor and male student (Carney et al. 2000).

The distribution of specific follow-up feedback in the second session almost exactly mirrored the distribution of specific goals; however, the proportion of specific feedback recorded could be increased. A fifth of students received only follow-up feedback of a generalised nature, often as non-descript praise. This kind of feedback does not contribute to students' further development, and can be embarrassing to students (Ende 1983). The students' evaluations of the second session also clearly supported specific follow-up feedback. In a fifth of cases, no follow-up feedback was documented,

but two-thirds of these students were still satisfied, and have possibly received appropriate verbal feedback.

The proportion of specific goals and specific feedback increased significantly between the first three and last three semesters, although we could not show a significant association with supervisors' length of experience of the tool. We interpreted these findings as an improved 'feedback culture' in practice, also among less experienced supervisors.

The face validity of the assessment was ascertained by involvement of supervisors in the adaptation of the assessment tool before implementation, and content validity was ensured by alignment to the goals of the attachment. In the estimation of concurrent validity, we found a low correlation with scores on the written exam; this was an expected finding, and in accordance with a recent systematic review of structured tools for direct observation (Kogan et al. 2009). However, the correlation increased if only students with the lowest performance scores were analysed, a result that could support the idea of supervisors' ability to accurately identify students with an inadequate knowledge base. Internal consistency reliability of the tool was very high, suggesting more of a global assessment, and also in line with previous literature (Kogan et al. 2009). As for the educational impact of the formative assessment, we could confirm satisfaction with the assessment strategy among most students and supervisors. Furthermore, the narrative comments of the follow-up feedback, aligned to personal goals, could be interpreted as students' development in clinical performance, albeit not objectively measured.

To the best of our knowledge, no previous study has reported on feedback, applied at two separate sessions in clinical training. Neither have we been able to find research designed to investigate the effects of long-term routine use of an assessment tool. Other strengths of our study are the face and content validity, the acceptability of the strategy, the application of interactive feedback methods and the attempt to form a conception of students' achievements in practice.

Some limitations of the assessment strategy and the study should be considered. First, we were unable to examine the inter-rater reliability of the assessment, as the tool was not used in each of students' patient encounters, in contrast to most research on work-based assessment. However, summarised, formal assessment twice during the attachment was a feasible strategy under the ideal conditions of continuity and a personal supervisor. As our main purpose for the assessment was the provision of feedback, we considered the sampling of cases and problems in general practice adequate for acceptable reliability in most cases, even if evaluations were more subjective (van der Vleuten & Schuwirth 2005). Second, despite recommendations on additional feedback on a day-to-day basis, some supervisors might have delayed the provision of feedback to a scheduled summarised session. It is widely acknowledged that for maximum effect informal feedback should take place shortly after an event, while the event is still fresh in mind (Cantillon & Sargeant 2008). However, in supporting students' development of more advanced integrated professional skills, delayed feedback can be advantageous (Hattie & Timperley 2007; Archer 2010). Third, we were unable to perform an evaluation of supervisors' benefits of the workshops on effective feedback,

as participation was voluntary; less than a third of supervisors attended, and nearly half of these participated after one or two semesters' experience with the tool.

The introduction of the tool emphasised the importance of direct observation, assessment and feedback in a clinical environment. After implementation of the tool, students' end-of-clerkship evaluations confirmed increased satisfaction with the feedback process, and the results from a follow-up investigation of the educational environment, as measured by DREEM, were significantly improved with regard to supervisors' provision of constructive criticism (Edgren et al. 2010).

Since the time of the study, the assessment tool has been slightly modified, and is now well established as documentary evidence for students' portfolio reflections (Haffling et al. 2010). Accordingly, it is now possible to examine cases of poor feedback and deficient student-supervisor relationships, which we could not accomplish during the study, due to the retrospective design.

Conclusions

This study extended our knowledge about work-based assessment and feedback by presenting a feasible strategy, applied to a continuous attachment in general practice with personal supervisors. The strategy was approved by students and supervisors. Almost all students were assessed and provided with feedback in the first session, most students were given specific goals, and a majority received specific follow-up feedback, aligned to their goals. Long-term experience of the assessment tool could successfully contribute to an increased stringency of assessment and an enhanced 'feedback culture' among supervisors. However, there was room for further improvement, and additional supervisor training in the provision of effective follow-up feedback was required.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on contributors

ANN-CHRISTIN HAFFLING, MD, PhD, general practitioner, working half-time as a university teacher in charge of the programme in Community Medicine, Lund University.

ANDERS BECKMAN, MD, PhD, general practitioner, part-time teacher on the programme in Community Medicine, Lund University.

GUDRUN EDGREN, PhD, MMedEd, senior lecturer and director of the Centre for Teaching and Learning of the Faculty of Medicine, Lund University

References

Archer JC. 2010. State of the science in health professional education: Effective feedback. *Med Educ* 44:101–108.

Braend AM, Frandsen Gran S, Frich JC, Lindbaek M. 2010. Medical students' clinical performance in general practice: Triangulating assessments from patients, teachers and students. *Med Teach* 32:333–339.

Burch VC, Seggie JL, Gary NE. 2006. Formative assessment promotes learning in undergraduate clinical clerkships. *S Afr Med J* 96:430–433.

Cantillon P, Sargeant J. 2008. Giving feedback in clinical settings. *BMJ* 337:1292–1294.

Carney PA, Dietrich AJ, Eliassen S, Pipas C, Donahue D. 2000. Differences in ambulatory teaching and learning by gender match of preceptors and students. *Fam Med* 32:618–623.

Colthart I, Bagnall G, Evans A, Allbutt H, Haig A, Illing J, McKinstry B. 2008. The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10. *Med Teach* 30:124–145.

Crabtree BF, Miller WL, editors. 1999. *Doing qualitative research*. Thousand Oaks, CA: Sage Publishing.

Daelmans HEM, Hoogenboom RJI, Donker AJM, Scherpbier AJJA, Stehouwer CDA, van der Vleuten CPM. 2004. Effectiveness of clinical rotations as a learning environment for achieving competences. *Med Teach* 26:305–312.

Daelmans HEM, Overmeer RM, van der Hem-Stokroos HH, Scherpbier AJJA, Stehouwer CDA, van der Vleuten CPM. 2006. In-training assessment: Qualitative study of effects on supervision and feedback in an undergraduate clinical rotation. *Med Educ* 40:51–58.

Edgren G, Haffling AC, Jakobsson U, McAleer S, Danielsen N. 2010. Comparing the educational environment (as measured by DREEM) at two different stages of curriculum reform. *Med Teach* 32:e233–e238.

Ende J. 1983. Feedback in clinical medical education. *JAMA* 250:777–781.

Eva KW, Regehr G. 2005. Self assessment in the health professions: A reformulation and research agenda. *Acad Med* 80:S46–S54.

Fernando N, Cleland J, McKenzie H, Cassar K. 2008. Identifying the factors that determine feedback given to undergraduate medical students following formative mini-CEX assessments. *Med Educ* 42:89–95.

Haffling AC, Beckman A, Pahlmblad A, Edgren G. 2010. Students' reflections in a portfolio pilot: Highlighting professional issues. *Med Teach* 32:e532–e540.

Hastings AM, McKinley RK, Fraser RC. 2006. Strengths and weaknesses in the consultation skills of senior medical students: Identification, enhancement and curricular change. *Med Educ* 40:437–443.

Hattie J, Timperley H. 2007. The power of feedback. *Rev Educ Res* 77:81–112.

Hewson MG, Little ML. 1998. Giving feedback in medical education. *J Gen Intern Med* 13:111–116.

Holmboe ES, Yepes M, Williams F, Huot SJ. 2004. Feedback and the mini clinical evaluation exercise. *J Gen Intern Med* 19:558–561.

Howley LD, Wilson WG. 2004. Direct observation of students during clerkship rotations: A multiyear descriptive study. *Acad Med* 79:276–280.

Kogan JR, Hauer KE. 2006. Brief report: Use of the mini-clinical evaluation exercise in internal medicine core clerkships. *J Gen Intern Med* 21:501–502.

Kogan JR, Holmboe ES, Hauer KE. 2009. Tools for direct observation and assessment of clinical skills of medical trainees. A systematic review. *JAMA* 302:1316–1326.

Kogan JR, Shea JA. 2008. Implementing feedback cards in core clerkships. *Med Educ* 42:1071–1079.

Lane JL, Gottlieb RP. 2000. Structured clinical observations: A method to teach clinical skills with limited time and financial resources. *Pediatrics* 105:973–977.

Maas C, Snijders TAB. 2003. The multi-level approach to repeated measures for complete and incomplete data. *Qual Quantity* 37:71–89.

McKinley RK, Fraser RC, van der Vleuten C, Hastings AM. 2000. Formative assessment of the consultation performance of medical students in the setting of general practice using a modified version of the leicester assessment package. *Med Educ* 34:573–579.

Norcini J, Burch V. 2007. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach* 29:855–871.

Paukert JL, Richards ML, Olney C. 2002. An encounter card system for increasing feedback to students. *Am J Surg* 183:300–304.

Pulito AR, Donnelly MB, Plymale M, Mentzer RM. 2006. What do faculty observe of medical students' clinical performance? *Teach Learn Med* 18:99–104.

Rashbash J, Charlton C, Browne WJ, Healy M, Cameron B. 2009. MiWin, Centre for Multilevel Modelling, University of Bristol.

Russell G. 2009. Taking time to watch. Observation and learning in family practice. *Can Fam Physician* 55:948–950.

Sargeant J, Armon H, Chesluk B, Dornan T, Eva K, Holmboe E, Lockyer J, Loney E, Mann K, van der Vleuten C. 2010. The processes and dimensions of informed self-assessment: A conceptual model. *Acad Med* 85:1212–1220.

Sargeant JM, Mann KV, van der Vleuten CP, Metsemakers JF. 2009. Reflection: A link between receiving and using assessment feedback. *Adv Health Sci Educ* 14:399–410.

Shumway JM, Harden RM. 2003. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Med Teach* 25:569–584.

van der Vleuten CPM. 1996. The assessment of professional competence: Developments, research and practical implications. *Adv Health Sci Educ* 1:41–67.

van der Vleuten CPM, Schuwirth LWT. 2005. Assessing professional competence: From methods to programmes. *Med Educ* 39:309–317.

Appendix

Assessment form

	1	2	3	4	5	6	7
<i>Medical interview</i>							
Welcomes patient, introduces him/herself to patient							
Demonstrates interest and respect towards patient							
Starts with an open question; encourages patient to continue							
Listens attentively to patient's opening statement, without interrupting or directing patient's response							
Uses pauses and silence appropriately							
Seeks clarification of words used by patient, if needed							
Recognises patient's verbal and non-verbal cues							
Phrases questions simply and clearly							
Expresses verbal empathy and support; appropriate non-verbal behaviour							
Elicits patients' ideas, concerns and expectations							
Summarises periodically and at the end to establish patient's reasons for consultation							
Negotiates agenda with patient, taking both patient's and doctor's needs into account							
<i>History taking</i>							
Elicits relevant and specific information from patients	1	2	3	4	5	6	7
Considers physical, social and psychological factors as appropriate							
<i>Physical examination</i>							
Performs physical examination sensitively, with respect for patients' integrity	1	2	3	4	5	6	7
Uses accurate examination technique							
Elicits physical signs correctly							
<i>Working diagnoses</i>							
Generates appropriate working diagnoses or identifies patient's problems after history taking	1	2	3	4	5	6	7
Seeks relevant and discriminating physical signs in examination to help confirm or refute working diagnoses							
<i>Problem solving</i>							
Correctly interprets and applies information, obtained from patient's records, history, physical examination and investigation	1	2	3	4	5	6	7
Is capable of applying knowledge of basic behavioural and clinical sciences to the identification and management of patient's problems							
Is capable of recognising limits of personal competence and acting accordingly							
<i>Investigations and treatment</i>							
Suggests investigations and/or treatment appropriate to findings and in collaboration with patients, taking into consideration discriminating use of resources	1	2	3	4	5	6	7
<i>Explanation and planning</i>							
Supplies patient with correct and clear information	1	2	3	4	5	6	7
Relates explanations to patient's perspective, aiming to achieve shared understanding							
Uses clear and understandable language, avoiding medical jargon							
Checks patient's understanding							
Arranges appropriate follow-up in dialogue with patient							
<i>Relationship with patients; time management</i>							
Maintains friendly but professional relationship with patient	1	2	3	4	5	6	7
Is capable of reflecting on common ethical issues in medical care							
Uses the available time for the consultation appropriately							