Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: informahealthcare.com/journals/imte20

The development of an instrument for evaluating clinical teachers: involving stakeholders to determine content validity

Renée E. Stalmeijer, Diana H.J.M. Dolmans, Ineke H.A.P. Wolfhagen, Arno M.M. Muijtjens & Albert J.J.A. Scherpbier

To cite this article: Renée E. Stalmeijer, Diana H.J.M. Dolmans, Ineke H.A.P. Wolfhagen, Arno M.M. Muijtjens & Albert J.J.A. Scherpbier (2008) The development of an instrument for evaluating clinical teachers: involving stakeholders to determine content validity, Medical Teacher, 30:8, e272-e277, DOI: 10.1080/01421590802258904

To link to this article: https://doi.org/10.1080/01421590802258904

Published online: 03 Jul 2009.

2008; 30: e272–e277 MEDICAL TEACHER

WEB PAPER

The development of an instrument for evaluating clinical teachers: involving stakeholders to determine content validity

RENÉE E. STALMEIJER[1], DIANA H.J.M. DOLMANS[1], INEKE H.A.P. WOLFHAGEN[1], ARNO M.M. MUIJTJENS[1] & ALBERT J.J.A. SCHERPBIER[2]

[1]Department of Educational Development and Research, Faculty of Health, Medicine, and Life Sciences, Maastricht University, The Netherlands, [2]Institute for Medical Education, Faculty of Health, Medicine, and Life Sciences, Maastricht University, The Netherlands

Abstract

Background: Research indicates that the quality of supervision strongly influences the learning of medical students in clinical practice. Clinical teachers need feedback to improve their supervisory skills. The available instruments either lack a clear theoretical framework or are not suitable for providing feedback to individual teachers. We developed an evaluation instrument based on the 'cognitive apprenticeship model'.

Aim: The aim was to estimate the content validity of the developed instrument.

Method: Item relevance was rated on a five-point scale (1 = highly irrelevant, 5 = highly relevant) by three groups of stakeholders in undergraduate clinical teaching: educationalists (N = 12), doctors (N = 16) and students (N = 12). Additionally, stakeholders commented on content, wording and omission of items.

Results: The items were generally rated as very relevant (Mean = 4.3, SD = 0.38, response = 95%) and any differences between the stakeholder groups were small. The results led to elimination of 4 items, rewording of 13 items and addition of 1 item.

Discussion: The cognitive apprenticeship model appears to offer a useful framework for the development of an evaluation instrument aimed at providing feedback to individual clinical teachers on the quality of student supervision. Further studies in larger populations will have to establish the instrument's statistical validity and generalizability.

Introduction

Students spend the final years of undergraduate medical education for the most part in clinical rotations in hospital settings, offering a potentially powerful learning environment because of its high authenticity and opportunities for active participation in clinical work and integrated learning of history taking, physical examination, clinical decision making and professionalism (Spencer 2003). However, the core activity of hospitals is patient care rather than clinical teaching. Time pressure, competing demands on staff from service, research, administration and teaching conspire to make the hospital a highly unstructured and complex learning environment (Spencer 2003). This means that students' learning experiences are largely determined by day-to-day events in the workplace rather than by pedagogical considerations (Collins et al. 1989). Research has shown that good supervision is the key to successful learning in clinical practice (Dolmans et al. 2002). Supervision can be defined as 'the provision of monitoring, guidance and feedback on matters of personal, professional, and educational development in the context of the doctor's care of patients' (Kilminster & Jolly 2000). The fact that supervision is a relatively infrequent occurrence in the clinical workplace (Grant et al. 2006) can be explained by the above-mentioned time pressures and competing demands on doctors but also by most doctors' lack of formal training for their role as a clinical teacher (Cottrell et al. 2002).

Practice points

- The cognitive apprenticeship model appears to offer a useful framework for designing an evaluation instrument aimed at providing feedback to clinical teachers on their supervision of medical students.
- The clinical supervisor evaluation instrument is considered relevant by three groups of stakeholders in undergraduate clinical training (educationalists, doctors and students).
- The positive rating of the instrument's content validity by three stakeholder groups suggests that the instrument will meet with broad acceptance when it is implemented in clinical training.
- Future research should focus on further validation of the instrument, on effective ways of providing feedback to clinical teachers based on student evaluations and on promoting faculty development based on student evaluations.

Correspondence: Renée Stalmeijer, Department of Educational Development and Research, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Tel: (+31) 43-3885735; fax: (+31) 43-3885779; email: r.stalmeijer@educ.unimaas.nl

ISSN 0142–159X print/ISSN 1466–187X online/08/080272–6 © 2008 Informa UK Ltd.
DOI: 10.1080/01421590802258904

Doctors need appropriate tools to aid them in selecting and developing effective ways of supervising students so as to create an effective learning environment. Appropriate evaluation of clinical teaching may guide doctors in this. Several instruments have been developed to evaluate clinical teachers' performance. In their 2004 review, Beckman and colleagues (Beckman et al. 2004) listed 21 articles describing instruments for providing feedback to doctors based on evaluations by undergraduate and graduate students. One of the most commonly used instruments was developed at Stanford University School of Medicine (Litzelman et al. 1998b) and contains categories based on educational and psychological theories of learning environments and empirical observations of clinical teaching. Although this is a validated instrument, its focus on different dimensions of teaching effectiveness in various settings, including small group sessions, makes it less suitable for specific feedback on individual supervision in the clinical setting. Another widely cited instrument, the Cleveland Clinic's Clinical Teaching Effectiveness Instrument (CCCTEI) (Copeland & Hewson 2000), was developed in cooperation with many stakeholders and tested for validity and reliability, but the theoretical dimensions underlying the items of this instrument are not clearly specified. Bowden and Marton (2000) observe that, for evaluation results to be helpful to teachers in optimizing their teaching practice, the evaluation instrument should have solid theoretical underpinnings. When evaluations are not grounded in theories of effective teaching, it is difficult to establish in which direction efforts to improve teaching should be headed and, consequently, real improvement is likely doubtful (Bowden & Marton 2000).

Medical education has always strongly relied on the 'traditional apprenticeship model' of clinical training, where students mainly learn by observing doctors as role models in day-to-day clinical practice. Although there is no denying that role modelling is a powerful way of transmitting values, attitudes and patterns of thinking and behaviour to students (Elzubeir & Rizk 2001), research indicates that effective learning depends strongly on active involvement of students and deliberate attention to cognitive processes underlying task performance. In line with these ideas, Collins et al. (1989) developed the concept of cognitive apprenticeship by rethinking the traditional apprenticeship model. The focus of the cognitive apprenticeship model is on experts' (tacit) cognitive processes when performing complex tasks. In this model, these processes are made explicit and strategies are given to help students observe, enact and practise with support from their teachers. The following methods are proposed for use by teachers to provide guided learning experiences to students.

- *Modelling*: students observe their teachers as they demonstrate certain parts of a task. Teachers should externalize their thinking processes as they perform the task by thinking aloud and explaining their judgement and reasoning to students.
- *Coaching*: teachers observe and help students in performing a task or in learning a task. The teacher offers feedback, scaffolding and modelling during the process.
- *Scaffolding*: teachers provide support and selective help to students in performing tasks they have not yet mastered. Teachers diagnose students' skill levels and stimulate students to move beyond their current level, which may mean that students need help. Eventually, as students' mastery advances, teachers reduce the level of support (fading).
- *Articulation*: teachers ask students to articulate their knowledge, reasoning or problem-solving processes.
- *Reflection*: teachers stimulate students to reflect on their own problem-solving processes and compare them with those of experts and other students.
- *Exploration*: students are pushed into a mode of problem solving on their own. Teachers formulate general goals for students and encourage them to focus on particular subgoals of interest.

According to research, another prerequisite for successful learning in the clinical environment, apart from the above methods, is a positive learning climate (Kilminster & Jolly 2000). Its importance in maximizing learning outcomes has also been emphasized by other studies. A way for teachers to create a propitious learning climate is to show that they respect their students and are interested in the students' learning (Litzelman et al. 1998b; Beckman et al. 2003).

We developed an instrument based on the cognitive apprenticeship model to elicit students' evaluations of individual doctors' clinical teaching to be used for feedback and eventually, improvement of the quality of teaching performance. In order to investigate the content validity of the instrument, we invited stakeholders from different settings to judge the instrument on relevance and wording. Quality being a relative concept (Harvey & Green 1993), it is only to be expected that individual experts will take different views of the quality of clinical teaching depending on their personal backgrounds. Therefore, we sought the opinions of different groups of stakeholders. We envisaged that this would help us not only to enhance the content validity of the instrument but also to broaden acceptance and support for the eventual implementation of the instrument (Guba & Lincoln 1989).

We approached three groups of stakeholders: doctors, educationalists and senior medical students, asking them to complete a questionnaire about the relevance of the instrument. The aim was to investigate and improve the instrument's content validity by addressing the following research question:

- How do three groups of stakeholders rate the relevance and wording of an instrument developed to be used by students to evaluate individual doctors' clinical teaching?

Method

Participants

Three groups of stakeholders were approached: 12 educationalists, 16 doctors and 12 senior medical students. We chose these groups for their combined theoretical knowledge

(educationalists) and practical experience in the domain of clinical teaching (doctors and students). The educationalists were from the eight Dutch medical schools and had expertise in medical education, especially clinical teaching and clerkships. The doctors were from various disciplines (internal medicine, paediatrics, surgery, obstetrics/gynaecology, ENT and dermatology) and were involved in clerkship teaching in different hospitals in the Netherlands. All doctors had had ample experience in supervising medical students and some were clerkship coordinators. Student representatives at Maastricht Medical School recruited student volunteers who had completed at least 6 of the total of 12 clinical rotations and have had experience with being supervised in clinical practice. As soon as 12 volunteers had come forward, recruitment was stopped. Participation was voluntary, participants received no reward and the data were anonymized. In Dutch medical schools, it is not required to submit this type of study to the ethics committee for approval.

Instrument

We developed a questionnaire about the relevance and quality of the items of the evaluation instrument. The items reflect the teaching methods of the cognitive apprenticeship model (Collins et al. 1989): modelling (four items), coaching (four items), scaffolding (four items), articulation (four items), reflection (four items), exploration (four items), general learning climate (three items) and miscellaneous (three items) (Table 1). Items were derived from the extensive description of the teaching methods in the cognitive apprenticeship model as provided by Collins et al. (1989). Respondents were asked to rate item relevance on a five-point Likert scale (1 = highly irrelevant; 5 = highly relevant). When they rated an item as 1 or 2, stakeholders were asked to give a written explanation in the space provided under the item in question. They could use the same space to write comments about item quality and wording. Finally, the participants were asked to indicate whether they had noticed any omissions and/or would like to add an item or items to the questionnaire.

Procedure

The questionnaire was sent by post with a return envelope and a letter informing the stakeholders of the purpose of the study. In the letter, it was explained that they were invited to participate because of their particular expertise in clinical teaching. After 3 weeks, a reminder was emailed to the non-responders with the questionnaire attached and requesting them to complete and return the questionnaire online.

Analysis

We used SPSS 14.0 to analyse the data. First, we checked for outliers and non-normality. Means and SDs of the relevance ratings were calculated by item, by teaching method and by stakeholder group. Because the data distribution was skewed, we used non-parametric analysis to examine the differences between the stakeholder groups in their ratings of the teaching methods. We determined frequencies, cross tabulations (ratings of relevance per teaching method for three groups of stakeholders) and Pearson chi-square. Additionally, Bonferroni correction was performed. Finally, effect sizes were calculated based on the formula described by Hojat and Xu (2004) to estimate effect sizes for chi-square:

$$\text{Effect size} = \sqrt{C^2/(1 - C^2)},$$

where C is '(...) the coefficient of contingency, which is a widely used measure of association between discrete measures in contingency tables, which can be derived from χ^2'. Stakeholders' comments on the items they considered less relevant were collected in a word file and analysed to see if the item should be retained, illuminated or reworded.

Criteria for item inclusion/exclusion

We eliminated from the questionnaire all the items that were rated below 3.5 by all three stakeholder groups. This prevented automatic exclusion of items receiving low ratings from one stakeholder group only. When an item gave rise to strong negative comments about the quality of item focus or wording, it was also eliminated.

Results

Response

In total, 10 out of 12 educationalists responded (83%), all 16 doctors responded (100%) and all 12 senior medical students responded (100%).

Descriptive statistics

Generally, item relevance is rated highly by all three stakeholder groups, with ratings varying between 3.4 and 4.6 (Table 1). The stakeholder groups' mean ratings of the relevance of the teaching methods are shown in Table 1. The mean overall ratings of the relevance of the teaching methods range from 3.8 to 4.5.

Pearson chi-square and effect sizes

After Bonferroni correction, Pearson chi-square analysis revealed significant differences between the stakeholder groups for four teaching methods (Table 2). Because of the non-normality of the data, several columns of the cross tabulations were collated in order to meet the preconditions for Pearson chi-square. For 'modelling', 'scaffolding', 'reflection' and 'exploration', points 1 and 2 of the five-point scale were collated, leaving four dimensions. For 'coaching', 'articulation' and 'miscellaneous', we collated 1, 2 and 3 into one column, which left three scoring dimensions. 'General learning climate' was excluded from further analysis because, even after 1, 2 and 3 were collated, the cross tabulation showed empty cells and thus did not satisfy the preconditions for Pearson chi-square. The results of the Pearson chi-square analyses of the three stakeholder groups are shown in Table 2.

Table 1. Mean ratings and SDs, on a Likert scale (1 = highly irrelevant; 5 = highly relevant), of the relevance of the items relating to eight supervision methods by stakeholder group and for the three stakeholder groups combined and overall group ratings per supervision method.

	Educationalists		Doctors		Students		Total item score
Supervision methods	M	SD	M	SD	M	SD	M
Modelling							
The clinical teacher:							
1. demonstrated how different tasks should be performed.	4.5	0.53	4.5	0.63	3.6	0.80	4.2
2. explained, while performing a task, which aspects were important and why.	4.4	0.84	4.7	0.48	4.2	1.00	4.4
3. created sufficient opportunities for me to observe him or her.	4.4	0.70	4.4	0.51	3.8	0.87	4.2
4. was a role model for me.	4.3	1.10	4.3	0.68	3.2	1.00	3.9
Overall ratings on modelling	4.4	0.41	4.5	0.45	3.7	0.47	
Coaching							
The clinical teacher:							
5. observed me while I was performing a task.	4.9	0.32	4.5	0.63	4.2	0.83	4.5
6. provided me with constructive and concrete feedback during or following direct observation.	4.6	0.70	4.7	0.48	4.5	0.70	4.6
7.* provided me with constructive and concrete feedback during my rotation.	4.3	1.10	4.5	0.90	4.7	0.50	4.5
8. gave me a better insight into aspects of my performance that needed improvement.	4.3	0.80	4.6	0.63	4.0	0.85	4.3
Overall ratings on coaching	4.5	0.50	4.6	0.60	4.3	0.50	
Scaffolding							
The clinical teacher:							
9. adjusted his/her teaching activities to my level of experience and competence.	4.1	1.00	4.2	0.75	3.8	0.62	4.0
10. allowed me to perform tasks that fit my level of experience and competence.	4.3	1.10	4.4	0.63	4.2	1.10	4.3
11. was supportive when I experienced difficulties with a task.	4.0	1.20	4.5	0.73	4.1	0.51	4.2
12. gradually decreased the amount of guidance in order to bolster my independence.	3.4	1.10	4.5	0.52	4.3	0.75	4.1
Overall ratings on scaffolding	4.0	0.66	4.4	0.47	4.1	0.44	
Articulation							
The clinical teacher:							
13. asked me to explain my reasoning and arguments.	4.6	0.70	4.6	0.63	4.5	0.52	4.6
14. alerted me to gaps in my knowledge and skills.	4.2	0.63	4.4	0.62	3.9	1.10	4.2
15. asked questions to increase my understanding.	4.3	0.71	4.6	0.62	4.5	0.70	4.5
16. stimulated me to ask questions to increase my understanding.	4.5	0.71	4.6	0.50	4.0	1.10	4.4
Overall ratings on articulation	4.0	0.50	4.5	0.48	4.4	0.57	
Reflection							
The clinical teacher:							
17. stimulated me to think about my own strengths and weaknesses.	4.4	0.84	4.4	0.62	3.4	0.90	4.1
18. stimulated me to think about how to improve my own strengths and weaknesses.	4.5	0.71	4.6	0.63	3.4	1.10	4.2
19.* stimulated me to compare my own approach to that of a more experienced doctor.	3.9	1.20	3.7	1.10	2.8	0.87	3.4
20.* stimulated me to reflect on the profession of a medical doctor.	3.5	0.85	3.8	1.00	3.0	1.30	3.4
Overall ratings on reflection	4.4	0.66	4.1	0.70	3.1	0.63	
Exploration							
The clinical teacher:							
21. stimulated me to formulate my own goals.	4.7	0.50	4.1	0.85	3.5	0.67	4.1
22. stimulated me to achieve my own goals.	4.8	0.42	4.3	0.45	3.8	0.75	4.2
23. challenged me to explore new tasks and possibilities.	4.5	0.85	3.9	0.72	4.3	0.75	4.2
24.* challenged me to extend my boundaries.	3.6	1.50	3.8	0.83	4.1	0.90	3.8
Overall ratings on exploration	4.5	0.41	4.0	0.59	3.9	0.54	
General Learning Climate							
The clinical teacher:							
25. established a safe-learning environment.	4.4	0.84	4.8	0.45	4.1	0.67	4.4
26. showed an interest in me as a student.	4.4	0.70	4.7	0.48	4.2	0.83	4.4
27. treated me with respect.	4.7	0.50	4.9	0.34	4.3	0.62	4.6
** *took enough time to supervise me*.							
Overall ratings on general learning climate	4.3	0.50	4.8	0.38	4.2	0.56	
Miscellaneous							
28. Give an overall mark (1–10) for this doctor as a clinical teacher.	3.7	1.60	4.2	0.83	3.6	1.20	3.9
29. What are the strengths of this clinical teacher? (open-ended question)	4.7	0.49	4.6	0.62	4.4	0.52	4.6
30. Which aspects of the performance of this clinical teacher can be improved? (open-ended question)	4.7	0.52	4.8	0.40	4.3	0.67	4.6
Overall ratings on miscellaneous	4.3	0.85	4.5	0.36	4.1	0.49	

*= Eliminated item; ** = added item.

Table 2. Pearson chi-square and effect sizes (ratings of relevance per teaching method for three groups of stakeholders).

Teaching method	χ^2	df	Effect size
Modelling	33.521*	6	0.47
Coaching	8.276	4	–
Scaffolding	23.004*	6	0.39
Articulation	5.162	4	–
Reflection	27.123*	6	0.42
Exploration	28.255*	6	0.43
General learning climate[a]	–	–	–
Miscellaneous	6.806	4	–

[a]Preconditions for Pearson chi-square not met.
*Bonferonni corrected significance level $p < 0.01$.

'Coaching', 'articulation' and 'miscellaneous' show no significant differences between the three groups. Compared to the other groups, the students give lower ratings for 'modelling' and 'reflection' and the doctors and the educationalists give higher ratings for 'scaffolding' and 'exploration', respectively. The effect sizes of the significant between-group differences are medium-to-large, based on an interpretation of effect sizes for chi-square of around 0.10, 0.30 and 0.50 as indicative of negligible practical importance, moderate practical importance and crucial practical importance, respectively (Cohen 1987; Hojat & Xu 2004) (Table 2).

Qualitative suggestions

The number of stakeholders' comments on the teaching methods varies between 7 and 18 per method. Comments pertain to the relevance and wording of items and most are from the educationalists. We cite some of the stakeholders' comments on the content of teaching methods to illustrate how qualitative considerations, in addition to the quantitative criterion led to elimination of items. The numbers refer to individual participants.

Coaching. Two of the items relating to 'coaching' focus on the importance of feedback (items 6 and 7). It is mainly educationalists who observe overlap of these items and point out that it is important for feedback to be given immediately after observation of the activity in question.

> (...) *I think that feedback should be provided as quickly as possible after a presentation in order for the feedback to be effective.* (Educationalist 9.10)

Reflection. All three groups of stakeholders comment on items 19 and 20. A doctor and a student highlight the risk of a counterproductive effect of item 19.

> *With this aspect one risks counterproductive results. When you compare yourself to more experienced doctors you can end up discouraging yourself, especially in a learning climate that is not safe.* (Doctor 1.16)
>
> (...) *Of course the approach of an experienced doctor differs from that of a student, that is self-evident, and this comparison does not give you much.* (Student 1.12)

Respondents from each group comment that item 20 is formulated too broadly.

> *It is not clear to me what is meant exactly. Is this an overarching question about what it means to be a doctor?* (Student 12.12)

Exploration. Item 24 is critiqued by educationalists and doctors for its suspected potential to stimulate irresponsible behaviour.

> (...) *I think this item is too 'wild' and dangerous. Possibly stimulates irresponsible behaviour.* (Educationalist 3.10)
>
> *I think* (...) *item 24 depends on the individual. Ninety percent of students already extend their boundaries and when a very enthusiastic supervisor comes along and adds new (boundary crossing) tasks, it may be counterproductive to the learning process.* (Doctor 15.16)

Modifications of the questionnaire

Items 19 and 20 were removed in response to low ratings by all stakeholders. Item 7 was removed because of overlap with item 6, which is retained as the more important item because of its focus on feedback immediately after or during observation. Concerns about incitement to irresponsible behaviour led to the removal of item 24. The wording of several items was altered and one item was added to 'General Learning Climate': 'The clinical teacher took enough time to supervise me'.

Discussion and conclusions

Hospitals are potentially very powerful learning environments for medical students (Spencer 2003). To fully realize this potential, high-quality supervision of students is of the essence and this can be improved when clinical teachers are evaluated and provided with feedback (Dolmans et al. 2002). As part of the development of a valid and reliable instrument to evaluate clinical teachers, we asked three groups of stakeholders in undergraduate clinical training to assess the content validity of an instrument based on the cognitive apprenticeship model of Collins et al. (1989).

The generally high ratings of the relevance of questionnaire items by all the stakeholders not only support the content validity of the instrument but also bode well for its broad acceptance when it is implemented (Guba & Lincoln 1989). As a consequence of both the qualitative and the quantitative results of this study, 4 of the total of 30 items were eliminated, 13 were modified and 1 was added. All three stakeholder groups contributed to modifications of the questionnaire. Of two 'coaching' items on feedback, which overlapped according to educationalists, the one item referring to the importance of immediate feedback after or during observed task performance was retained.

Although students gave low ratings for reflection and commented that they considered stimulation of reflection more a task for mentors than for clinical supervisors, their ratings were offset by higher ratings from educators and doctors. This was not the case with students' low ratings of 'reflection' items about comparing student performance with that of more experienced doctors and reflecting on the profession of a medical doctor. An additional reason for removing these items were stakeholders' comments that these items were too vague and general. Concern expressed by doctors and educationalists about potential harmful effects of encouraging students to extend their boundaries was the decisive factor in the removal of item 24.

The results of this study are encouraging in that they support and helped improve the content validity of the instrument based on the cognitive apprenticeship model which we developed to evaluate clinical teachers' performance. Future studies will have to address the statistical validation of the instrument and the generalizability of results to determine the number of students needed to provide reliable feedback to doctors about their supervisory skills. After statistical validation of the instrument, further studies should investigate whether the feedback-based outcomes of the evaluation instrument is beneficial to clinical teachers and can help them improve their teaching performance. Because research has demonstrated that feedback from students on clinical teachers' performance can have mixed effects on teaching effectiveness (Litzelman et al. 1998a), it is also important to investigate how feedback from student evaluations is best communicated to clinical teachers to facilitate improvement of teaching effectiveness. The role of faculty development programmes in this respect should also be addressed.

Notes on contributors

RENÉE E STALMEIJER, MSc, is an Educationalist and Researcher. She works at the Department of Educational Development and Research at the Faculty of Health, Medicine and Life Sciences at Maastricht University, the Netherlands.

DIANA HJM DOLMANS, PhD, is an Associate Professor and Educational Psychologist. She works at the Department of Educational Development and Research at the Faculty of Health, Medicine and Life Sciences at Maastricht University, the Netherlands.

INEKE HAP WOLFHAGEN, PhD, is an Associate Professor and Educational Psychologist. She works at the Department of Educational Development and Research at the Faculty of Health, Medicine, and Life Sciences at Maastricht University, the Netherlands.

ARNO MM MUIJTJENS, PhD, is an Assistant Professor, Statistician and Methodologist at the Department of Educational Development and Research at the Faculty of Health, Medicine, and Life Sciences at Maastricht University, the Netherlands.

ALBERT JJA SCHERPBIER, MD PhD, is a Professor of Medical Education and Scientific Director of the Institute for Medical Education, Faculty of Health, Medicine, and Life Sciences, Maastricht University, the Netherlands.

Acknowledgements

The authors would like to thank all the stakeholders for their feedback on the instrument. They also wish to thank Mereke Gorsira for editing the final version.

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

References

Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. 2004. How reliable are assessments of clinical teaching? A review of the published instruments. J Gen Intern Med 19:971–977.

Beckman TJ, Lee MC, Rohren CH, Pankratz VS. 2003. Evaluating an instrument for the peer review of inpatient teaching. Med Teach 25:131–135.

Bowden J, Marton F. 2000. The University of Learning. London: Kogan Page Limited.

Cohen J. 1987. Statistical Power Analysis for Behavioral Sciences. Hillsdale, NJ: Erlbaum.

Collins A, Brown JS, Newman SE. 1989. Cognitive apprenticeship: teaching the crafts of reading, writing, and mathematics. In: Resnick LB, editor. Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Copeland HL, Hewson MG. 2000. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical centre. Acad Med 75:161–166.

Cottrell D, Kilminster S, Jolly B, Grant J. 2002. What is effective supervision and how does it happen? A critical incident study. Med Educ 36:1042–1049.

Dolmans DHJM, Wolfhagen HAP, Essed GGM, Scherpbier AJJA, Van der Vleuten CPM. 2002. Students' perceptions of relationships between some educational variables in the out-patient setting. Med Educ 36:735–741.

Elzubeir MA, Rizk DEE. 2001. Identifying characteristics that students, interns and residents look for in their role models. Med Educ 35: 272–277.

Grant J, Kilminster S, Jolly B, Cottrell D. 2006. Clinical supervision of SpRs: Where does it happen, when does it happen and is it effective? Med Educ 37:140–148.

Guba EG, Lincoln YS. 1989. Fourth Generation Evaluation. London, Newbury Park: Sage Publications.

Harvey L, Green D. 1993. Defining quality. Assess Eval Higher Educ 18:9–34.

Hojat M, Xu G. 2004. A visitor's guide to effect sizes. Statistical significance versus practical (clinical) importance of research findings. Adv Health Sci Educ 9:241–249.

Kilminster SM, Jolly BC. 2000. Effective supervision in clinical practice settings: A literature review. Med Educ 34:827–840.

Litzelman DK, Stratos GA, Marriott DJ, Lazaridis EN, Skeff KM. 1998a. Beneficial and harmful effects of augmented feedback on physicians' clinical-teaching performances. Acad Med 73:324–332.

Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. 1998b. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. Acad Med 73:688–695.

Spencer J. 2003. Learning and teaching in the clinical environment. ABC of learning and teaching in medicine. Br Med J 326:591–594.