

REVIEWS



Assessing the Quality of Clinical Teachers

A Systematic Review of Content and Quality of Questionnaires for Assessing Clinical Teachers

Cornelia R. M. G. Fluit, MD, MEdSci¹, Sanneke Bolhuis, PhD¹, Richard Grol, PhD², Roland Laan, MD, PhD³, and Michel Wensing, PhD²

¹Department for Evaluation, Quality and Development of Medical Education, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; ²Scientific Institute for Quality of Healthcare, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; ³Department of Rheumatology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

BACKGROUND: Learning in a clinical environment differs from formal educational settings and provides specific challenges for clinicians who are teachers. Instruments that reflect these challenges are needed to identify the strengths and weaknesses of clinical teachers.

OBJECTIVE: To systematically review the content, validity, and aims of questionnaires used to assess clinical teachers.

DATA SOURCES: MEDLINE, EMBASE, PsycINFO and ERIC from 1976 up to March 2010.

REVIEW METHODS: The searches revealed 54 papers on 32 instruments. Data from these papers were documented by independent researchers, using a structured format that included content of the instrument, validation methods, aims of the instrument, and its setting.

RESULTS: Aspects covered by the instruments predominantly concerned the use of teaching strategies (included in 30 instruments), supporter role (29), role modeling (27), and feedback (26). Providing opportunities for clinical learning activities was included in 13 instruments. Most studies referred to literature on good clinical teaching, although they failed to provide a clear description of what constitutes a good clinical teacher. Instrument length varied from 1 to 58 items. Except for two instruments, all had to be completed by clerks/residents. Instruments served to provide formative feedback (instruments) but were also used for resource allocation, promotion, and annual performance review (14 instruments). All but two studies reported on internal consistency and/or reliability; other aspects of validity were examined less frequently.

CONCLUSIONS: No instrument covered all relevant aspects of clinical teaching comprehensively. Validation of the instruments was often limited to assessment of internal consistency and reliability. Available instruments for assessing clinical teachers should be used carefully, especially for consequential decisions. There is a need for more valid comprehensive instruments.

Electronic supplementary material The online version of this article (doi:10.1007/s11606-010-1458-y) contains supplementary material, which is available to authorized users.

Received October 29, 2009

Revised February 22, 2010

Accepted July 2, 2010

Published online August 12, 2010

KEY WORDS: medical education–assessment methods; medical education–faculty development; medical student and residency education; systematic reviews; clinical teaching–instruments.

J Gen Intern Med 25(12):1337–45

DOI: 10.1007/s11606-010-1458-y

© Society of General Internal Medicine 2010. This article is published with open access at Springerlink.com

BACKGROUND

High-quality patient care is only feasible if physicians have received high-quality teaching during both their undergraduate and their residential years.^{1,2} Their medical development starts in a university environment and continues in a clinical setting, where they predominantly learn on the job. Most teaching in clinical settings is provided by physicians who also work in that clinical setting. Therefore, it is important that these physicians should be good and effective teachers.^{3,4}

There is a considerable body of literature on the roles of clinical teachers, including several review studies.^{5,10} Excellent clinical teachers are described as physician role models, effective supervisors, dynamic teachers, and supportive individuals, possibly complemented by their role as assessors, planners, and resource developers.^{11,13} Some of this literature, including a recent review, has described good clinical teachers by looking for typical behaviors or characteristics, which often fit into one or more of the above-mentioned roles.^{14,18} There is also a considerable body of literature defining physicians in terms of single roles, such as role models or supervisors.^{3,19,25} Effective role models are clinically competent, possess excellent teaching skills, and have personal qualities, such as compassion, sense of humor and integrity. Effective supervisors give feedback and provide guidance, involve their students in patient care, and provide opportunities for carrying out procedures. Studies on work-based learning show that work allocation and structuring are important for learners to make progress and that a significant proportion of their work needs to be sufficiently new to challenge them without being so daunting as to reduce their confidence.^{26,29} To assign work that provides effective learning opportunities, therefore, is essential.³⁰

Physician competencies which should be acquired by trainees during their training have recently been formulated.^{1,4,31,32} Clinical teachers should at least role model these competencies.³³ Box 1 summarizes the roles of the clinical teacher.

DOMAINS OF CLINICAL TEACHING	
<p>Physician role model the clinical teacher can be observed during daily work</p> <p>Teacher role the clinical teacher uses effective teaching strategies such as discussing learning goals, giving explanations, asking questions, discussing work, and giving instructions for further learning</p> <p>Supervisor role Assigning work effective for learning: the clinical teacher makes sure that trainees perform tasks they can learn from, participate in daily practice with growing responsibility and sufficient autonomy Providing feedback: the clinical teacher in his or her supervisor role provides feedback in order to improve performance and stimulate the learning process (formative use)</p>	<p>Supportive person the clinical teacher contributes to a positive, stimulating learning environment by being supportive, enthusiastic, friendly, accessible</p> <p>Assessor role the clinical teacher assesses the performance of a trainee, using different assessment tools, in order to make go/no go decisions (summative use)</p> <p>Planner/organizer the clinical teacher plans teaching moments in daily work and takes time for the trainee and education</p> <p>Resource developer the clinical teacher develops educational materials</p>
MEDICAL COMPETENCIES	
<p>Medical Expert (eg. demonstrate diagnostic and therapeutic skills for ethical and effective patient care and demonstrate effective consultation services with respect to patient care, education and legal opinions)</p> <p>Communicator (eg. establish therapeutic relationships with patients/families and discuss appropriate information with patients/families and the health care team)</p> <p>Collaborator (eg. consult effectively with other physicians and health care professionals and contribute effectively to other interdisciplinary team activities)</p>	<p>Manager (eg. utilize resources effectively to balance patient care, learning needs, and outside activities and work effectively and efficiently in a health care organization)</p> <p>Health Care Advocate (eg. identify the important determinants of health affecting patients and contribute effectively to improved health of patients and communities)</p> <p>Scholar (eg. develop, implement and monitor a personal continuing education strategy and critically appraise sources of medical information)</p> <p>Professional (eg. deliver highest quality care with integrity, honesty and compassion and exhibit appropriate personal and interpersonal professional behaviors)</p>

Box 1. Domains of clinical teaching.

The assessment of clinical teachers in postgraduate education is often based on questionnaires completed by residents.²⁵ It is important that these instruments should have good measurement properties. If used to help improve clinical teaching skills, such instruments should provide reliable and relevant feedback on clinical teachers' strengths and weaknesses.^{6,25} If used for promotion and tenure, or ranking of clinical teachers, instruments should be able to distinguish between good and bad teachers in a highly valid and reliable way.

The American Psychological and Education Research Associations published standards identifying five sources of validity evidence by: (1) Content, (2) Response process, (3) Internal structure, (4) Relations to other variables, and (5) Consequences (see Box 2).^{34,37}

Beckman³⁴ extensively reviewed instruments for their psychometric qualities, thereby giving useful recommendations on ways of improving this quality. However, we are unaware of any

studies that focus specifically on the content of these questionnaires in relation to literature on good clinical teaching and on how instruments are used in practice. Therefore, we performed a review of instruments for assessing clinical teachers in order to determine (1) the content of these instruments (what they measure) and (2) how well these instruments measure clinical teaching (their construction and use).

METHODS

Identification of Studies

We searched the MEDLINE, EMBASE, PsycINFO, and ERIC databases from 1976 through March 2010 (see online appen-

Validity source evidence	Definition	Examples
Content	The relationship between a test's content and the construct it is intended to measure. Refers to themes and wording of items. Includes experts' input. Also included development strategies to ensure appropriate content representation	-Representativeness of items to domain -Test specifications -Quality of test questions
Response process	Analyses of responses, including strategies and thought processes of individual respondents. Differences in response processes may reveal sources of variance that are irrelevant to the construct being measured. Also includes instrument security, scoring, and reporting of results	-Student format familiarity -Quality control of items -Costs -Minimum number of respondents needed
Internal structure	The degree to which items fit underlying construct. Most often reported as measures of internal consistency and factor analysis	-Item analysis -Inter-item correlations -Score scale reliability -Standard errors of measurement: -generalizability -dimensionality -factor analysis
Relation to other variables	The relationship between scores and other variables relevant to the construct being measured. Relationships may be positive (convergent or predictive) or negative (divergent or discriminant)	-Correlation with other variables -Generalizability of evidence
Consequences	Surveys are intended to have some desired effect, but they also have unintended effects. Evaluating such consequences can support or challenge the validity or score interpretations	Formative or summative use: -Impact of scores on learners -Consequences on learning -Go/no go decision -False positives/negatives
Sources: Downing ^{36,37} , Beckman ³⁴ , Boor ³⁵		

Box 2. Five sources of validity evidence.

dix Search Strategy). Search terms included clinical teaching, clinical teacher, medical teacher, medical education, evaluation, effectiveness, behavior, instrument, and validity. A manual search was performed by reviewing references of retrieved articles and contents of medical education journals. Two authors (CF and SB) independently reviewed the titles and abstracts of retrieved publications for possible inclusion in the review. If the article or the instrument was not available (N=5), we contacted the author(s). Studies were included after they had been reviewed by two authors (CF and SB) to make sure that they: (1) reported on the development, validation, or application of an instrument for measuring clinical teacher performance; (2) contained a description of the content of the instrument; (3) were applied in a clinical setting (hospital or primary care); (4) used clerks, residents, or peers for assessing clinical teachers. We restricted our review to studies published in English.

Data Extraction

A standardized data extraction form was developed and piloted to abstract information from the included articles. Data extraction was done by three authors: the first author (CF) assessed all selected articles; two other authors (SB and MW) each assessed half the articles. Disagreements about data extraction merely concerned the content of six questionnaires, which were discussed by the three authors until consensus was reached.

The content of the instruments was assessed in two ways. First, we ascertained to what degree these instruments reflected the domains described in the literature as being characteristic of good clinical teaching (see Box 1). As the instruments we

examined focused on teaching in daily clinical practice, we excluded the domain of 'resource developer' as this concerns an off-the-job activity which, moreover, may be more difficult for trainees to assess. Secondly, we wanted to know to what extent instruments assessed the way clinical teachers teach their residents the medical competencies of physicians. Therefore, we used the medical competencies as described by the Canadian Medical Educational Directives (CanMEDS): medical expert, communicator, collaborator, manager, health advocate, scholar, and professional, as these have been widely adopted. Good physician educators would be expected to act as role models of these competencies and be effective teachers of these competencies.^{4,33,38}

The five sources for validity evidence served to analyze the psychometric qualities of the instruments (Box 2). Information was extracted about the study population, the setting where the instrument was used, evaluators, number and type of items, feasibility of the instrument (duration, costs, and number of questionnaires needed), the aim of the instrument, and how the instrument had been developed.

Reliability coefficients estimate measurement error in assessing and quantifying the measurement's consistency.³⁷ The most frequently used estimates are Cronbach's α (based on the test-retest concept and indicating internal consistency), the Kappa statistic (a correlation coefficient indicating inter-rater reliability), ANOVA (also indicating inter-rater statistics), and generalizability theory (to estimate the concurrent effect of multiple-source reliability, note that this not refer to external validity of a measure). Comparisons with other instruments or related variables were documented.

Finally, the purpose of the instrument was documented: feedback (formative assessment) or promotion and tenure (summative assessment).

RESULTS

We found 2,712 potentially relevant abstracts, 155 of which were retrieved for full text review (see online Appendix Flow Diagram). Application of the inclusion criteria resulted in 54 articles.^{33,38,90} As some articles were about the same instrument, a total of 32 instruments was found. Table 1 presents their general characteristics. Instruments were most frequently used in an inpatient clinical setting ($N=20$) and tested in one

discipline ($N=25$). Instruments were completed by residents ($N=16$), students/clerks ($N=18$), trained observers ($N=2$), or peers ($N=1$). Most instruments ($N=28$) were developed in the USA. There was a wide range in the number of teachers (9-711, median 41) and evaluators (2-731, median 66) involved in validation of the instruments.

Content of the Questionnaires

The 'teacher' and 'supporter' domains were represented most frequently in the instruments (30 and 29 instruments), followed by 'role model' (27 instruments), and 'feedback' in 26 instruments (see online Appendix Table A). Together, these were expressed by 479 (79%) of all items. Most of these items concerned teaching techniques (216 items, 36%). The domain of 'planning teaching activities' was represented by 33 items in 18

Table 1. Characteristics of the Instruments for Measuring Clinical Teachers

Author	First publ. Ref.no	Add. Publ. Ref.no	Instrument	Setting	Disciplines	Teachers (N)	Evaluators (N)	Evaluators (type)	Country
Afonso 2005	39	-		I	IntM,CCU	30	83	S,R	USA
Beckman 2003	40	41,42,43	MTEF	I	IntM	10	3	P	USA
Bergen 1993	44	-	CTORS	I	IntM	40	-	TO	USA
Cohen 1996	50	73	TES	I	Surg	43	-	S,R,F	USA
Copeland 2000	51	45,46,47, 48,49,88	CTEI	I/O	IntM/Ped/ Surg Anes/ Path/Radiol	711	-	S,R,F	USA
Cox 2001	52	-	-	I/OR	Surg	20	49	R	USA
de Oliveira 2008	53	-	-	I	Anes	39	19	R	Brasil
Dolmans 2004	54	-	-	I	Ped	13	-	C	Netherlands
Donnelly 1989	55	-	-	?	IntM	300	100	C	USA
Donner-Banzhof 2003	56	-	-	GP	GP	80	80	R	Germany
Guyatt 1993	57	-	-	I/A	IntM	41	-	C,I,R	USA
Hayward 1995	58	-	-	O	IntM	15	-	R	USA
Hekelman 1993	59	60	-	GP	GP	16	2	TO	USA
Hewson 1990	61	-	WICT	I	IntM	9	28	R	USA
Irby 1981	62	63,75,76	CTAF	I	Gyn	230	320	S	USA
James 2002	64	65	MedEdIQ	O	IntM/Ped/ GP	156	131	C	USA
Lewis 1990	66	-	-	GP	GP	10	24	R	USA
Litzelman 1998	67	68,69,71	SFDP	W	GenM	178	374	C	USA
Love 1982	70	-	-	Pharmacy	IntM/Ped/ Surg ED/ AmbC	39	66	C	USA
McLeod 1991	72	-	-	-	IntM	35	50	S	USA
Mullan 1993	74	-	-	I	Ped	-	-	C	USA
Schum 1993	79	78	OTE	I	Ped	186	375	S,C,R	USA
Shellenberger 1982	80	-	PEQ	GP	GP	-	-	C	USA
Silber 2006	38	-	-	I	IntM/Surg	11	57	R	USA
Smith 2004	33	-	-	I	IntM	99	731	R	USA
Solomon 1997	81	-	-	I/O	GenM	147	-	C	USA
Spickard 1996	82	-	-	I	IntM	44	91	C	USA
Stalmeijer 2008	84	83	-	-	All	-	-	C	Netherlands
Steiner 2000	85	86	-	I	ED	29	-	R	USA
Tortolani 1991	87	77	-	I	Surg	62	23	R	USA
Williams 1998	89	-	-	I	IntM	203	29	R	USA
Zuberi 2007	90	-	SETOC	O	IntM/Ped/ Surg/Gyn/ GP/Opt/Oto	87	224	C	USA/ Pakistan

instruments. 'Assigning work that is effective for learning' was represented by 29 items, that is 5% of all items, in 13 instruments. Items about 'assessment' were represented by nine items (2% of all items) in five instruments. Fifteen instruments asked for overall teaching quality or effectiveness (OTE). Seven instruments contained one question or several questions that were either open questions or questions that were not directly related to the quality of the individual teacher (other).

About one-third of all items (213 items) could be related to the competencies as described by CanMEDS (see online Appendix Table B). The other items did not refer specifically to competencies as specified in the CanMEDS roles. More than half of these (129 items) were related to the medical expert (102 items) and scholarship (27 items) competencies, evaluating the teaching of medical skills and knowledge (e.g., 'the physician showed me how to take a history'; 'uses relevant scientific literature in supporting his/her clinical advice'). There were 42 items on professional behavior. Role modeling and teaching (101 and 71 items) are strategies most frequently associated with the teaching and learning of competencies (e.g., 'is a role model of conscientious care', 'sympathetic and considerate towards patients').

Measurement Characteristics of the Instruments

The measurement characteristics of the validated instruments have been summarized in Table 2. The content validity of most instruments was based on the literature and the input of experts and residents/students. In 17 studies, a previously developed instrument served as a basis for the development of the new instrument. Irby's questionnaire (1986) was mentioned four times for the development of an instrument.^{39,70,72,79} Not all studies documented what previously developed instrument had been used. Five studies reported the use of a learning theory for questionnaire construction.^{54,65,67,84,90} The number of available evaluations varied from 30 to 8,048 (median 506). Instruments contained 1 to 58 items, with Likert scale points ranging from four to nine. Information about feasibility in terms of costs, time needed for filling in the questionnaire, or minimum number of questionnaires needed was reported in eight instruments.

Studies represented a variety of validity evidence procedures, with the most common one being the determination of internal consistency by internal structure by factor analysis and/or Cronbach's α (20 studies). Less common validation methods were determining inter-rater and intra-class correlations, Pearson correlation coefficients, Spearman Brown formula, and studies using the statistical generalizability theory. In some studies, scores were compared to the overall teaching score or scores on other instruments. Some studies reported on hypotheses formulated in advance or compared scores of different respondent groups.^{33,43,46,55,64,70}

The reported purposes of clinical teacher evaluations are summarized in Table C (see online Appendix Table C). Not all authors documented how their instrument was to be used. Although providing feedback is the evaluation aim mentioned most frequently, 14 authors reported that the instrument was or would be used for summative purposes such as promotion, tenure, or resource allocation.

CONCLUSION

Our review revealed 32 instruments designed for evaluating clinical teachers. These instruments differ in terms of content and/or quality of the measurement.

What do the Instruments (Not) Measure?

Most instruments cover the important domains of teaching, role modeling, supporting, and providing feedback, roles that have been emphasized in the literature on clinical teaching.

Items on assessment of residents are least represented in the instruments. Assessment is becoming more and more important since society is increasingly demanding accountability from its doctors.⁹¹ With the shift towards competency-based residents' training programs, there is also a growing need for measuring competency levels and competency development, including not only knowledge and skills but also performance in practice.^{1,91,93} For all these reasons, assessing residents by using a mix of instruments is an important task in clinical teaching.^{93,96}

Items on the supervisor's role in assigning clinical work and planning are also under-represented in the instruments. Opportunities for participating in the clinical work environment and for performing clinical activities are crucial for residents' development.³⁰ Planning in the demanding clinical environment provides structure and context for both teachers and trainees, as well as a framework for reflection and evaluation.⁹⁷ Creating and safeguarding opportunities for performing relevant activities and planning teaching activities can therefore be seen as key evaluable roles of clinical teachers.

Teaching and learning in the clinical environment need to focus on relevant content. Doctors' competencies in their roles as medical experts, professionals, and scholars were well represented in the instruments, but doctors' competencies in their roles as communicators, collaborators, health advocates, and managers were less frequently measured. We found two instruments that reflected all CanMEDS competencies.^{38,61} Remarkably, one instrument had been developed in 1990, before the CanMEDS roles were published.⁶¹

In summary, although all instruments cover important parts of clinical teaching, no instrument covers all clinical teaching domains. Therefore, the use of any of these individual assessment tools will be limited.

How Well Do the Instruments Measure Clinical Teaching?

Construction of the Questionnaires. The inpatient setting was used most frequently for validating and/or applying instruments, and most instruments were used in only one discipline (most frequently internal medicine). These limitations restrict the generalizability of the instruments. Different teaching skills may be required for instruction in outpatient versus inpatient settings.^{72,75,98} Some authors found no differences in teaching behavior in relation to the setting.⁵² However, Beckman compared teaching assessment scores of general internists and cardiologists and found factor instability, thus highlighting the importance of validating assessments for the specific contexts in which they are used.⁴³

Table 2. Measurement Characteristics of Instruments for Measuring Clinical Teachers

instrument	Validity source evidence						Internal structure	Relation to other variables
	Content		Response process			Feasibility		
	Source of items	Content validity	Evaluations (N)	Items (N)	Likert scale			
Afonso	1	1,2,4	199	18	5	-	FA Cronbach's α	Anonymous and open evaluations compared
Beckman	1,2,3	1,2,3,4	30	28	5	1,2,3	Cronbach's α Kendall's Tau	Follow-up compared scores of residents and peers
Bergen	1,2	1,2,3,4	-	21	5	-	Interrater agreement	-
Cohen	-	1,2	3750	4	5	3	ICC	-
Copeland	1,2,3	1,2,3,4,5,6	8048	15	5	-	FA Cronbach's α G-coefficient	Follow-up compared scores of residents and peers Compared to OTS
Cox	2,4	1,2,4,5	753	20	5	-	Cronbach's α	-
de Oliveira	4	1,2,4,5,6	954	11	4	-	Cronbach's α Inter Item C G-Study	Compared to overall perception of quality
Dolmans	2,4,5	1,3,4,5,6	-	18	5	-	-	-
Donnelly	-	1,2	952	12	7	-	-	Hypotheses formulated in advance
Donner-Banzhof	1,2,3,4	1,2,4,5	80	41	4	-	Cronbach's α Pearson r	-
Guyatt	2,4	-	-	14	5	1	FA Intra domain correlation	-
Hayward	1,4	1,2,3,4,5,6	142	18	5	3	FA Cronbach's α G-Study	-
Hekelman	2,4	2,3,5,6	160	17	-	-	Cronbach's α ICC	-
Hewson	2,4	1,2,4,5	-	46	5	-	Cronbach's α	-
Irby	1,2	1,2,3,4	1567	9	5	3	Spearman Brown Pearson r	Qualitative comparison with other instruments
James	2,4,5	1,2,5	156	58	6	-	FA Cronbach's α	Scores compared with grades of students
Lewis	1,2,4	1,2,4,5	-	16	VAS	-	ICC interrater correlation	Compared qualitative evaluation data
Litzelman	1,2,5	1,2,3,4	1581	25	5	1	FA Cronbach's α Inter Item C	-
Love	1,2	1,2,3,4	281	9	5	1,3	Pearson r	Residents and attendings compared
McLeod	1,4	1,4,5	-	25	6	-	FA Kruskal-Wallis Wilcoxon Rank	-
Mullan	4	1,4,5	-	23	-	-	Standardized alpha	Compared to OTS
Schum	1,2	1,2,4	749	10	7	-	FA Cronbach's α	Compared to OTS
Shellenberger	1,2,4	1,2,4,5	-	34	4	-	FA Cronbach's α	-
Silber	4	1,2,4,5,6	226	22	5	-	Product-moment correlation	-
Smith	2,4	1,2,3,4,5,6	731	32	5	1,3	Cronbach's α Inter Item C Inter-rater reliability SEM	Hypotheses formulated in advance/ Scores compared with grades of students
Solomon	-	1	2185	13	4	-	ICC Spearman Brown SEM Inter-rater reliability	-
Spickard	1	1,2,3	-	9	9	-	FA Cronbach's α	-
Stalmeijer	1,2,3,5	1,2,3,4,5,6	-	30	-	-	-	-
Steiner	1	1,2	48	4	5	-	-	Compared with other instruments
Tortolani	-	1,2	-	10	5	-	FA Pearson r	-
Williams	-	1,2	203	1	5	-	ICC	Correlation with leadership
Zuberi	1,2,3,5	1,2,3,4,5,6	-	15	7	-	Cronbach's α Inter Item C G-coefficient	ROC curves

Source of items

1 = previously developed instrument, 2 = analysis of literature, 3 = observations, 4 = expert opinions,

5 = learning theory

Content validity

1 = measurement aim described, 2 = target population described, 3 = clear framework of overall concept, 4 = item construction described, 5 = target population involved, 6 = items tested

Feasibility

Information provided about: 1 = duration of the test, 2 = costs, 3 = minimum number of respondents needed

Abbreviations

VAS = Visual Analog Scale, FA = Factor analysis, ICC = Intra Class Correlation, Inter Item C = Inter Item Correlation, G-Study = Generalizability study, SEM = Standard Error of Mean, ROC = Receiver Operating Characteristic curve, OTS = overall teaching score

Most authors used factor analysis and Cronbach's α to demonstrate an instrument's dimensionality and internal consistency, respectively. Less commonly used methods included the establishment of validity by showing convergence between new and existing instruments, and by correlating faculty assessments with educationally relevant outcomes. Computing Cronbach's α or completing a factor analysis may be the simplest statistical analysis to carry out for rating-scale data, but these analyses do not provide sufficient validity evidence for rating scales.⁹⁹ Validity is a unified concept and should be approached as a hypothesis, requiring that multiple sources of validity evidence be gathered to support or refute the hypothesis.^{36,99,100} This suggests that a broader variety of validity evidence should be considered when planning clinical teacher assessments.

Use of the Questionnaires. As most instruments are completed by residents and/or students, there are several issues that may affect ratings. First, residents tend to rate their teachers very highly,³⁴ which may cause ceiling effects, but this is rarely discussed in the selected studies. Second, learners at different stages differ in what they appreciate most in teachers and, hence, may rate their clinical teachers differently.^{9,39} Last, anonymous evaluations reveal lower scores than non-anonymous evaluations.^{9,39} Though most questionnaires are anonymous, anonymity may not be realistic in a department with only a few residents. Therefore, as part of an instrument's validation process, it should be tested in different settings, in different disciplines, by involving learners at different stages of the learning process, and by taking different evaluation circumstances into account. Even if the assessment of clinical teachers by residents could reveal valid information, evaluations should be derived from multiple and diverse sources, including peers and self-assessment, to allow "triangulation" of assessments.¹⁰¹

Limitations of this Study

As in any systematic review, we may have failed to identify instruments. Our search was limited to English-language journals, which may have introduced publication bias.

Implications

Instruments for assessing clinical teachers are used to provide feedback but also to back up consequential decisions relating to promotion, tenure, and resource allocation. In order to improve clinical teaching, assessments need to be effective in informing clinical teachers about all important domains and in identifying individual faculty strengths and weaknesses.^{33,66} Therefore, it is first of all important that the full assessment package includes all aspects of clinical teaching. In addition to the well known domains of teaching, role modeling, providing feedback, and being supportive, other domains also need attention, particularly the domains of assigning relevant clinical work, assessing residents, and planning teaching activities. Real improvement is more likely to be accomplished if all important domains are included in the selected set of assessment

instruments.^{33,54,84} This would likely require multiple complementary evaluation instruments.

Secondly, further study is needed to determine whether instruments can be validly used to assess a wider range of clinicians in different settings and different disciplines. Thirdly, evidence of an instrument's validity should be obtained from a variety of sources. Fourthly, we need to determine what factors influence evaluation outcomes, for instance, year of residency and non-anonymous versus anonymous evaluations. Finally, optimal assessment needs to balance requirements relating to measurement characteristics, content validity, feasibility, and acceptance.¹⁰² The primary requirement for any assessment, however, is that it should measure what it stands for, that is teaching in the clinical environment.

Acknowledgement: The authors would like to thank Rikkert Stuve of 'The Text Consultant' for editing of the manuscript.

This work was funded by the Department for Evaluation, Quality and Development of Medical Education of the Radboud University Nijmegen Medical Centre.

Conflict of Interest: None disclosed.

Open Access: This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Corresponding Author: Cornelia R. M. G. Fluit, MD, MEdSci; Department for Education, Quality and Development of Medical Education, Radboud University Nijmegen Medical Centre, 306 IWOO, PO Box 9101, 6500HB Nijmegen, The Netherlands (e-mail: C.Fluit@owi.umcn.nl).

REFERENCES

1. Leach DC. Changing education to improve patient care. *Qual Health Care.* 2001;10(Suppl 2):ii54-8.
2. Leach DC, Philibert I. High-quality learning for high-quality health care: getting it right. *JAMA.* 2006;296(9):1132-4.
3. Kilminster S, Cottrell D, Grant J, Jolly B. AMEE Guide No. 27: effective educational and clinical supervision. *Med Teach.* 2007;29(1):2-19.
4. Prideaux D, Alexander H, Bower A, et al. Clinical teaching: maintaining an educational role for doctors in the new health care environment. *Med Educ.* 2000;34(10):820-6.
5. Boor K, Teunissen PW, Scherpier AJ, van der Vleuten CP, van de Lande J, Scheele F. Residents' perceptions of the ideal clinical teacher - A qualitative study. *Eur J Obstet Gynecol Reprod Biol.* 2008 May 1.
6. Irby DM, Ramsey PG, Gillmore GM, Schaad D. Characteristics of effective clinical teachers of ambulatory care medicine. *Acad Med.* 1991;66(1):54-5.
7. Irby DM. Teaching and learning in ambulatory care settings: a thematic review of the literature. *Acad Med.* 1995;70(10):898-931.
8. Parsell G, Bligh J. Recent perspectives on clinical teaching. *Med Educ.* 2001;35(4):409-14.
9. Ullian JA, Bland CJ, Simpson DE. An alternative approach to defining the role of the clinical teacher. *Acad Med.* 1994;69(10):832-8.
10. Paukert JL, Richards BF. How medical students and residents describe the roles and characteristics of their influential clinical teachers. *Acad Med.* 2000;75(8):843-5.
11. Harden RM, Crosby J. AMEE Guide No 20: the good teacher is more than a lecturer - the twelve roles of the teacher. *Med Teach.* 2000;22(4):334-47.
12. Hesketh EA. A framework for developing excellence as a clinical educator. *Med Educ.* 2001;35(6):555-64.

13. **Ramani S, Leinster S.** AMEE Guide no. 34: teaching in the clinical environment. *Med Teach.* 2008;30(4):347-64.
14. **Bandiera G, Lee S, Tiberius R.** Creating effective learning in today's emergency departments: how accomplished teachers get it done. *Ann Emerg Med.* 2005;45(3):253-61.
15. **Bleakley A.** Pre-registration house officers and ward-based learning: a 'new apprenticeship' model. *Med Educ.* 2002;36(1):9-15.
16. **Chitsabesan P, Corbett S, Walker L, Spencer J, Barton JR.** Describing clinical teachers' characteristics and behaviours using critical incidents and repertory grids. *Med Educ.* 2006;40(7):645-53.
17. **Heidenreich C, Lye P, Simpson D, Lourich M.** The search for effective and efficient ambulatory teaching methods through the literature. *Pediatrics.* 2000;105(1 Pt 3):231-7.
18. **Sutkin G, Wagner E, Harris I, Schiffer R.** What makes a good clinical teacher in medicine? A review of the literature. *Acad Med.* 2008;83(5):452-66.
19. **Cote L, Leclere H.** How clinical teachers perceive the doctor-patient relationship and themselves as role models. *Acad Med.* 2000;75(11):1117-24.
20. **Kenny NP, Mann KV, MacLeod H.** Role modeling in physicians' professional formation: reconsidering an essential but untapped educational strategy. *Acad Med.* 2003;78(12):1203-10.
21. **Wright SM, Kern DE, Kolodner K, Howard DM, Brancati FL.** Attributes of excellent attending-physician role models. *N Engl J Med.* 1998;339(27):1986-93.
22. **Wright SM, Carrese JA.** Excellence in role modelling: insight and perspectives from the pros. *CMAJ.* 2002;167(6):638-43.
23. **Althouse LA, Stritter FT, Steiner BD.** Attitudes and approaches of influential role models in clinical education. *Adv Health Sci Educ Theory Pract.* 1999;4(2):111-22.
24. **Kilminster SM, Jolly BC.** Effective supervision in clinical practice settings: a literature review. *Med Educ.* 2000;34(10):827-40.
25. **Snell L, Tallett S, Haist S, et al.** A review of the evaluation of clinical teaching: new perspectives and challenges. *Med Educ.* 2000;34(10):862-70.
26. **Bolhuis S.** Professional development between teachers' practical knowledge and external demands: plea for a broad social-constructivist and critical approach. In: **Oser FK, Achtenwagen F, Renold U, eds.** *Competence Oriented Teacher Training.* Old Research Demands and New Pathways. Rotterdam: Sense Publishers; 2006:237-49.
27. **Cheetham G, Chivers G.** How professionals learn: an investigation of informal learning amongst people working in professions. *J Eur Ind Train.* 2001;25(5):248-92.
28. **Eraut M.** Informal learning in the workplace. *Stud Contin Educ.* 2004;26(2):247-73.
29. **Fluit C, Bolhuis S.** *Teaching and Learning. Patient-centered Acute Care Training (PACT).* 1 ed. European Society of Intensive Care Medicine; 2007.
30. **Teunissen PW, Scheele F, Scherpbier AJ, et al.** How residents learn: qualitative evidence for the pivotal role of clinical activities. *Med Educ.* 2007;41(8):763-70.
31. **Frank JR, Danoff D.** The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Med Teach.* 2007;29(7):642-7.
32. **Leach DC.** Evaluation of competency: an ACGME perspective. Accreditation council for graduate medical education. *Am J Phys Med Rehabil.* 2000;79(5):487-9.
33. **Smith CA, Varkey AB, Evans AT, Reilly BM.** Evaluating the performance of inpatient attending physicians: a new instrument for today's teaching hospitals. *J Gen Intern Med.* 2004;19(7):766-71.
34. **Beckman TJ, Cook DA, Mandrekar JN.** What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med.* 2005;20(12):1159-64.
35. **Boor K.** *The Clinical Learning Climate.* Amsterdam: Vrije Universiteit Amsterdam; 2009.
36. **Downing SM.** Validity: on meaningful interpretation of assessment data. *Med Educ.* 2003;37(9):830-7.
37. **Downing SM.** Reliability: on the reproducibility of assessment data. *Med Educ.* 2004;38(9):1006-12.
38. **Silber C, Novielli K, Paskin D, et al.** Use of critical incidents to develop a rating form for resident evaluation of faculty teaching. *Med Educ.* 2006;40(12):1201-8.
39. **Afonso NM, Cardozo LJ, Mascarenhas OA, Aranha AN, Shah C.** Are anonymous evaluations a better assessment of faculty teaching performance? A comparative analysis of open and anonymous evaluation processes. *Fam Med.* 2005;37(1):43-7.
40. **Beckman TJ, Lee MC, Rohren CH, Pankratz VS.** Evaluating an instrument for the peer review of inpatient teaching. *Med Teach.* 2003;25(2):131-5.
41. **Beckman TJ, Lee MC, Mandrekar JN.** A comparison of clinical teaching evaluations by resident and peer physicians. *Med Teach.* 2004;26(4):321-5.
42. **Beckman TJ, Mandrekar JN.** The interpersonal, cognitive and efficiency domains of clinical teaching: construct validity of a multi-dimensional scale. *Med Educ.* 2005;39(12):1221-9.
43. **Beckman TJ, Cook DA, Mandrekar JN.** Factor instability of clinical teaching assessment scores among general internists and cardiologists. *Med Educ.* 2006;40(12):1209-16.
44. **Bergen M, Stratos G, Berman J, Skeff KM.** Comparison of clinical teaching by residents and attending physicians in inpatient and lecture settings. *Teach Learn Med.* 1993;5:149-57.
45. **Bierer SB, Fishleder AJ, Dannefer E, Farrow N, Hull AL.** Psychometric properties of an instrument designed to measure the educational quality of graduate training programs. *Eval Health Prof.* 2004;27(4):410-24.
46. **Bierer SB, Hull AL.** Examination of a clinical teaching effectiveness instrument used for summative faculty assessment. *Eval Health Prof.* 2007;30(4):339-61.
47. **Brujin M, Busari JO, Wolf BH.** Quality of clinical supervision as perceived by specialist registrars in a university and district teaching hospital. *Med Educ.* 2006;40(10):1002-8.
48. **Busari JO, Weggelaar NM, Knottnerus AC, Greidanus PM, Scherpbier AJ.** How medical residents perceive the quality of supervision provided by attending doctors in the clinical setting. *Med Educ.* 2005;39(7):696-703.
49. **Busari JO, Koot BG.** Quality of clinical supervision as perceived by attending doctors in university and district teaching hospitals. *Med Educ.* 2007;41(10):957-64.
50. **Cohen R, MacRae H, Jamieson C.** Teaching effectiveness of surgeons. *Am J Surg.* 1996;171(6):612-4.
51. **Copeland HL, Hewson MG.** Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center. *Acad Med.* 2000;75(2):161-6.
52. **Cox SS, Swanson MS.** Identification of teaching excellence in operating room and clinic settings. *Am J Surg.* 2002;183(3):251-5.
53. **de Oliveira Filho GR, Dal Mago AJ, Garcia JH, Goldschmidt R.** An instrument designed for faculty supervision evaluation by anesthesia residents and its psychometric properties. *Anesth Analg.* 2008;107(4):1316-22.
54. **Dolmans DH, Wolfhagen HA, Gerver WJ, de Grave W, Scherpbier AJ.** Providing physicians with feedback on how they supervise students during patient contacts. *Med Teach.* 2004;26(5):409-14.
55. **Donnelly MB, Woolliscroft JO.** Evaluation of clinical instructors by third-year medical students. *Acad Med.* 1989;64(3):159-64.
56. **Donner-Banzhoff N, Merle H, Baum E, Basler HD.** Feedback for general practice trainers: developing and testing a standardised instrument using the importance-quality-score method. *Med Educ.* 2003;37(9):772-7.
57. **Guyatt G, Nishikawa J, Willan A, et al.** A measurement process for evaluating clinical teachers in internal medicine. *Can Med Assoc J.* 1993;149(8):1097-102.
58. **Hayward R.** Why the culture of medicine has to change. *BMJ.* 2007;335(7623):775.
59. **Hekelman F, Vanek E, Kelly K, Alemagno S.** Characteristics of family physicians' clinical teaching behaviors in the ambulatory setting: a descriptive study. *Teach Learn Med.* 1993;5(1):18-23.
60. **Hekelman F, Snyder C, Clint W, Alemagno S, Hull A, et al.** Humanistic teaching attributes of primary care physicians. *Teach Learn Med.* 1995;7(1):29-36.
61. **Hewson MG, Jensen NM.** An inventory to improve clinical teaching in the general internal medicine clinic. *Med Educ.* 1990;24(6):518-27.
62. **Irby DM, Rakestraw P.** Evaluating clinical teaching in medicine. *J Med Educ.* 1981;56:181-6.
63. **Irby DM, Gillmore GM, Ramsey PG.** Factors affecting ratings of clinical teachers by medical students and residents. *J Med Educ.* 1987;62(1):1-7.
64. **James PA, Osborne JW.** A measure of medical instructional quality in ambulatory settings: the MediQ. *Fam Med.* 1999;31(4):263-9.
65. **James PA, Kreiter CD, Shipengrover J, Crosson J.** Identifying the attributes of instructional quality in ambulatory teaching sites: a validation study of the MedEd IQ. *Fam Med.* 2002;34(4):268-73.

66. **Lewis BS, Pace WD.** Qualitative and quantitative methods for the assessment of clinical preceptors. *Fam Med.* 1990;22(5):356-60.
67. **Litzelman DK.** Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Acad Med.* 1998;73(6):688-95.
68. **Litzelman DK, Stratos GA, Marriott DJ, Lazaridis EN, Skeff KM.** Beneficial and harmful effects of augmented feedback on physicians' clinical-teaching performances. *Acad Med.* 1998;73(3):324-32.
69. **Litzelman DK, Westmoreland GR, Skeff KM, Stratos GA.** Factorial validation of an educational framework using residents' evaluations of clinician-educators. *Acad Med.* 1999;74(10 Suppl):S25-7.
70. **Love DW, Heller LE, Parker PF.** The use of student evaluations in examining clinical teaching in pharmacy. *Drug Intell Clin Pharm.* 1982;16(10):759-64.
71. **Marriott DJ, Litzelman DK.** Students' global assessments of clinical teachers: a reliable and valid measure of teaching effectiveness. *Acad Med.* 1998;73(10 Suppl):S72-4.
72. **McLeod PJ.** Faculty perspectives of a valid and reliable clinical tutor evaluation program. *Eval Health Prof.* 1991;14(3):333-42.
73. **Mourad O, Redelmeier DA.** Clinical teaching and clinical outcomes: teaching capability and its association with patient outcomes. *Med Educ.* 2006;40(7):637-44.
74. **Mullan P, Sullivan D, Dielman T.** What are raters rating? Predicting medical student, pediatric resident, and faculty ratings of clinical teachers. *Teach Learn Med.* 1993;5(4):221-6.
75. **Ramsbottom-Lucier MT, Gillmore GM, Irby DM, Ramsey PG.** Evaluation of clinical teaching by general internal medicine faculty in outpatient and inpatient settings. *Acad Med.* 1994;69(2):152-4.
76. **Ramsey PG, Gillmore GM, Irby DM.** Evaluating clinical teaching in the medicine clerkship: relationship of instructor experience and training setting to ratings of teaching effectiveness. *J Gen Intern Med.* 1988;3(4):351-5.
77. **Risucci DA, Lutsky L, Rosati RJ, Tortolani AJ.** Reliability and accuracy of resident evaluations of surgical faculty. *Eval Health Prof.* 1992;15(3):313-24.
78. **Schum TR, Yindra KJ.** Relationship between systematic feedback to faculty and ratings of clinical teaching. *Acad Med.* 1996;71(10):1100-2.
79. **Schum T, Koss R, Yindra K, Nelson D.** Students' and residents' ratings of teaching effectiveness in a department of pediatrics. *Teach Learn Med.* 1993;5(3):128-32.
80. **Shellenberger S, Mahan JM.** A factor analytic study of teaching in off-campus general practice clerkships. *Med Educ.* 1982;16(3):151-5.
81. **Solomon DJ, Speer AJ, Rosebraugh CJ, DiPette DJ.** The reliability of medical student ratings of clinical teaching. *Eval Health Prof.* 1997;20(3):343-52.
82. **Spickard A III, Corbett EC Jr, Schorling JB.** Improving residents' teaching skills and attitudes toward teaching. *J Gen Intern Med.* 1996;11(8):475-80.
83. **Stalmeijer RE, Dolmans DH, Wolfhagen IH, Scherpbier AJ.** Cognitive apprenticeship in clinical practice: can it stimulate learning in the opinion of students? *Adv Health Sci Educ Theory Pract.* 2008 September 17.
84. **Stalmeijer RE, Dolmans DH, Wolfhagen IH, Muijtjens AM, Scherpbier AJ.** The development of an instrument for evaluating clinical teachers: involving stakeholders to determine content validity. *Med Teach.* 2008;30(8):e272-7.
85. **Steiner IP, Franc-Law J, Kelly KD, Rowe BH.** Faculty evaluation by residents in an emergency medicine program: a new evaluation instrument. *Acad Emerg Med.* 2000;7(9):1015-21.
86. **Steiner IP, Yoon PW, Kelly KD, et al.** The influence of residents training level on their evaluation of clinical teaching faculty. *Teach Learn Med.* 2005;17(1):42-8.
87. **Tortolani AJ, Risucci DA, Rosati RJ.** Resident evaluation of surgical faculty. *J Surg Res.* 1991;51(3):186-91.
88. **van der Hem-Stokroos HH, van der Vleuten CPM, Daelmans HEM, Haarman HJTM, Scherpbier AJJA.** Reliability of the clinical teaching effectiveness instrument. *Med Educ.* 2005;39(9):904-10.
89. **Williams BC.** Validation of a global measure of faculty's clinical teaching performance. *Acad Med.* 2002;77(2):177-80.
90. **Zuberi RW, Bordage G, Norman GR.** Validation of the SETOC instrument - Student evaluation of teaching in outpatient clinics. *Adv Health Sci Educ Theory Pract.* 2007;12(1):55-69.
91. **Norcini JJ.** Current perspectives in assessment: the assessment of performance at work. *Med Educ.* 2005;39(9):880-9.
92. **Miller GE.** The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65(9 Suppl):S63-7.
93. **Norcini J, Burch V.** Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach.* 2007;29(9):855-71.
94. **Scheele F, Teunissen P, van Luijk S, et al.** Introducing competency-based postgraduate medical education in the Netherlands. *Med Teach.* 2008;30(3):248-53.
95. **ten Cate O.** Trust, competence, and the supervisor's role in postgraduate training. *BMJ.* 2006;333(7571):748-51.
96. **ten Cate O, Scheele F.** Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med.* 2007;82(6):542-7.
97. **Spencer J.** ABC of learning and teaching in medicine: learning and teaching in the clinical environment. *BMJ.* 2003;326(7389):591-4.
98. **McLeod PJ, James CA, Abrahamowicz M.** Clinical tutor evaluation: a 5-year study by students on an in-patient service and residents in an ambulatory care clinic. *Med Educ.* 1993;27(1):48-54.
99. **Downing SM, Haladyna TM.** Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ.* 2004;38(3):327-33.
100. **Auewarakul C, Downing SM, Jaturatamrong U, Praditsuwan R.** Sources of validity evidence for an internal medicine student evaluation system: an evaluative study of assessment methods. *Med Educ.* 2005;39(3):276-83.
101. **Jahangiri L, Mucciolo TW, Choi M, Spielman AI.** Assessment of teaching effectiveness in U.S. Dental schools and the value of triangulation. *J Dent Educ.* 2008;72(6):707-18.
102. **Van der Vleuten CPM.** The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract.* 1996;1(1):41-67.