WATCH: Warwick Assessment insulation for Clinical teaching: Development and testing

Sonia Ijaz Haider, Neil Johnson, Jill Elizabeth Thistlethwaite, Gay Fagan & Muhammad Furqan Bari


To link to this article: http://dx.doi.org/10.3109/0142159X.2014.947936

Published online: 26 Aug 2014.

Submit your article to this journal

Article views: 280

View related articles

View Crossmark data
WATCH: Warwick Assessment insTrument for Clinical teaChing: Development and testing

SONIA IJAZ HAIDER1,2, NEIL JOHNSON1, JILL ELIZABETH THISTLETHWAITE3, GAY FAGAN1 & MUHAMMAD FURQAN BARI4
1University of Warwick, UK, 2Shalamar Medical and Dental College, Pakistan, 3University of Queensland, Australia, 4Dow International Medical College, Pakistan

Abstract

Objective: Medical education and teaching skills are core competencies included in the generic curriculum for specialty training. To support the development of these skills, there is need for a validated instrument. This study aims to develop and test an instrument to measure the attributes of specialty trainees as effective teachers.

Methods: The study was conducted in two phases. In first phase, the content of the instrument was generated from the literature and tested using the Delphi technique. In second phase, the instrument was field tested for validity and reliability using factor analysis and generalizability study. Feasibility was calculated by the time taken to complete the instrument. Acceptability and educational impact were determined by qualitative analysis of written feedback. Attributes of specialty trainees were assessed by clinical supervisors, peers, and students.

Results: The Delphi study produced consensus on 15 statements which formed the basis of the instrument. In field study, a total of 415 instruments were completed. Factor analysis demonstrated a three-factor solution ('learning-teaching milieu', 'teaching skills', and 'learner-orientated'). A generalizability coefficient was 0.92. Mean time to complete the instrument was five minutes. Feedback indicated that it was an acceptable and useful method of assessment.

Conclusion: This new instrument provides valid, reliable, feasible, and acceptable assessment of clinical teaching.

Introduction

The significance of teaching in the medical profession can be traced back to the classical Hippocratic Oath (Coller et al. 2002). According to this oath, those who are experts in the field are expected to pass on their knowledge and skills to those who are junior to them and still considered as learners. The General Medical Council (GMC 1999, p. 1) emphasizes teaching by and among doctors.

Historically most of the formal teaching in medicine has been undertaken by senior doctors (consultants/specialists and general practitioners/family physicians) (Prideaux et al. 2000). A key reason is that they are at the highest level of experience and are, therefore, assumed to be competent to take on the role of teaching their more junior colleagues and students. However, with the increasing demands of patient care, administrative work, and other activities such as research, senior doctors have less time available to engage in frequent teaching activities (Seabrook 2003; Hendry et al. 2005). Consequently, there is an increasing emphasis on junior doctors in postgraduate training taking on formal teaching roles, particularly specialty trainees (also known as registrars and residents) because they are next in seniority and will be expected to teach effectively when they become consultants (specialists) in the future (Johnson et al. 1996; Bordley & LitzeIman 2000; Busari et al. 2003). It has been reported that specialty trainees are responsible for 80% of the educational activities of medical students and trainee doctors (Edwards et al. 2002). Specialty trainees are recognized as important and influential teachers because their attitudes and skills have been shown to influence learning (Barrow 1960; Busari & Scherphe 2004). It is, therefore, essential to assess the teaching ability of specialty trainees and a suitable instrument is required to do this.

Practice points

- WATCH is a 15-item instrument which can be used to assess clinical teaching among trainee doctors across various clinical specialties.
- It can be used for assessment by different assessors including clinical supervisors, peers and students.
- It is feasible to use it for observation of clinical teaching.
- The instrument demonstrates sound evidence of reliability and validity.
- Findings from the present study indicate that it is useful instrument which can be used to facilitate trainees in improving their clinical teaching.
Research on assessment of clinical teaching has outlined a set of guiding principles to broaden current perspectives on assessment (van der Vleuten & Schuwirth, 2005). In addition to demonstrating validity and reliability assessment, instruments should be feasible, acceptable to the stakeholders, easy to administer, and applicable to all levels of teachers (Snell et al. 2000; van der Vleuten & Schuwirth, 2005). A review (Fluit et al. 2010) of existing clinical teaching instruments concluded that none of them measures all the relevant aspects of clinical teaching and have focused on testing the psychometric properties (i.e., reliability and validity) (Beckman et al. 2005) only. Furthermore, the majority of the existing instruments have been designed for use by senior doctors (Fluit et al. 2010). While the attributes required by seniors and juniors doctors may be the same this cannot be assumed – for example, the content may not apply to junior doctors and the phrasing used when assessing senior doctors may require some change when assessments of trainees are undertaken. Any changes to the content of an instrument such as rewording existing items, removing items or adding new items would change those instruments and indicate that the properties of those instruments as previously measured would not necessarily apply. Rather than attempting to make adjustments to existing instruments or revalidate them, this study sets out to identify a priori, to develop and test an instrument for assessment specifically of clinical teaching among junior doctors.

Methods

Development of the instrument

The content of the instrument was developed by conducting a systematic literature review to identify attributes of effective clinical teachers. The search terms were ‘characteristics’, ‘attributes’, ‘teaching’, ‘clinical teachers’, ‘residents’, ‘specialty registrars’, and ‘trainees’. Medline, ERIC, Embase, PsycINFO, and EBSCOhost databases were searched from 1978 to 2012. All the articles exploring teaching in medicine and written in English were included. Articles not related to medicine or that were editorials and book reviews were excluded. A total of 46 articles were included in the review. A content analysis of these articles was performed to establish attributes identified recurrently as essential to clinical teaching. Sixteen such characteristics were identified.

To gain experts’ views on these attributes and establish their content validity, we undertook a consensus study. As our intended informants were spread over a wide geographical area, we used the Delphi approach (Gallagher et al. 1993; Murphy et al. 1998; Yousaf 2007) to seek the views of four groups of informants consisting of (a) medical directors (a physician who is usually employed by a hospital to serve in a medical and administrative capacity as head of the organized medical staff); (b) training programme directors (who are responsible for managing postgraduate specialty training programmes in the UK); (c) educational supervisors (who oversee the progress of individual trainees ensuring that they are making the necessary clinical and educational progress); and (d) specialty trainees. For this study, specialty trainees were defined as those trainees who had completed their foundation years training (years 1 and 2 following award of medical degree) and were specializing in a particular discipline of the medical profession, for example, internal medicine, surgery, or paediatrics. The rationale for choosing training programme directors and educational supervisors was that both these groups are directly involved in the teaching and learning of trainees; they are training experts themselves and have an in-depth understanding of the qualities required for specialty trainees to become effective teachers. Medical directors may not be directly involved in teaching but indirectly, as potential employers of trainees as teachers, they understand the requirements of potential future employees as teachers. Lastly, it was important to seek the opinions of those personnel for whom the instrument was being developed (Linstone & Turnoff 2002), therefore, we also included specialty trainees. Ethical approval was obtained from for the study.

There were three rounds of the Delphi study conducted online using Survey Monkey (www.surveymonkey.com). In round one, participants were asked to indicate their agreement or disagreement regarding the importance of each item. Participants were also given the option to modify the wording of items and list any additional characteristics which they considered were pertinent to clinical teaching. As reported in other similar Delphi studies (Boendermaker et al. 2003; Yeates et al. 2008), consensus was defined as 80% of participants indicating that they “agreed” or “strongly agreed” on the item as important for teaching. Qualitative analysis of participants’ comments was also undertaken. In round two participants were asked to rate the revised items because on the basis of qualitative and quantitative analysis of round one, items had been reworded, removed, and further items added. In round, three participants were asked to give their feedback, using free text, on the final suggested content of the instrument. All the participants were sent the results of each round of the Delphi study.

Testing the instrument

Between September 2011 and February 2012, trainees in all years and specialties involved in clinical teaching in the West Midlands region of the UK were sent a letter inviting them to participate in the study. The trainees were informed of the study procedures, and given the opportunity to ask any questions about the study. Participation in the study was voluntary and participants could withdraw from the study at any point without any repercussions.

Based on the existing evidence relating to reliability (Copeland & Hewson 2000), we asked each trainee to identify five assessors. Each specialty trainee was provided with five instruments, which they were asked to distribute to their assessors before undertaking a teaching session. Given the nature of the clinical setting, trainees were asked to choose their assessors from among senior doctors (consultant/general practitioner), peers or students; they were advised that they could invite all their assessors from the same group if necessary, for example, all peers or students. The assessors
were asked via formal written instructions to mark the teaching performance of the specialty trainee using the instrument at the end of the teaching session and to post their completed instrument in a prepaid sealed envelope to the chief investigator.

Reliability was estimated by conducting a generalizability analysis and computation of a G coefficient, using G strings IV (Bloch & Geoff 2011). As the G study required the same number of completed instruments for every specialty trainee, we randomly selected five instruments per specialty trainee (Copeland & Hewson 2000; de Oliveira Filho et al. 2008). For the study, we hypothesized that differences between assessors evaluating a trainee could come from the trainees, the assessors who were nested within the trainee (i.e., every trainee was assessed by different assessors) and the items. When computing the G coefficient, we chose a design where items were fixed at 15 (the final number of items included in the instrument) and assessors were random. A D-study (decision study) was also performed to estimate the G coefficient with different numbers of assessors. Construct validity was determined by factor analysis. This analysis allows the exploration of the underlying dimensions of the construct being assessed and indicates the variance accounted for by each factor (Lai et al. 2006). The feasibility of the instrument was determined by calculating the time taken to complete the instrument. Participants were also asked to indicate (free text) if they encountered any difficulty in completing the instrument. We explored the acceptability by analyzing open-ended comments to the following question: "In your opinion, is this teaching assessment an acceptable method of assessment or not?" Educational impact was also evaluated by analyzing responses to the following question, "In your opinion, can this teaching assessment have any impact on teaching by trainees?"

Results
Delphi results
A total of 112 (56%) participants responded in the first round of the Delphi study of whom 24 (21%) were medical directors, 32 (29%) training program directors, 29 (30%) educational supervisors, and 27 (24%) specialty trainees. The respondents came from various specialties, including medicine (16%), surgery (11%), gynecology and obstetrics (10%), ophthalmology (4%), radiology (6%), pathology (7%), psychiatry (10%), pediatrics (15%), anesthesia (12%), and general practice (9%). The mean teaching years' experience of medical directors was 9.1, of training programme directors was 12.5, and of educational supervisors was 11.2.

Table 1 presents the results of round one which indicate that except for one statement, "displays appropriate and relevant sense of humor," there was 80% or above consensus on all statements. Analysis of the comments indicated minor wording changes for some statements. Participants also recommended removing the statement "demonstrates adequate clinical and medical knowledge" because the statement "Is clinically competent (has sound analytical, diagnostic, therapeutic and reasoning skills)" was considered to indicate a similar attribute. Participants also advised that a statement about the use of aids and resources for teaching should be included.

There were 105 (52.5%) participants in the second round. Table 2 shows that there was consensus (80% or above) on all statements. Respondents did not provide any further comments or modification to the statements.

In the third round, respondents were presented with the final result of the statements to be included in the instrument. A total of 103 (51.5%) respondents participated in the third round. There was limited feedback on the final statements. Among the respondents who did provide feedback, examples of the comments were "this is fine," "nothing to add more,"

<table>
<thead>
<tr>
<th>Statements</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communicates effectively with students/trainees (listens attentively, answers clearly and explains logically with reason)</td>
<td>64.29</td>
<td>34.51</td>
<td>1.21</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Expresses enthusiasm towards teaching and learning</td>
<td>58.92</td>
<td>37.57</td>
<td>3.56</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Promotes active engagement of students during learning</td>
<td>26.85</td>
<td>64.28</td>
<td>7.24</td>
<td>1.62</td>
<td>0</td>
</tr>
<tr>
<td>Avoids favoritism, racism, negative criticism, and discrimination</td>
<td>58.57</td>
<td>36.14</td>
<td>3.71</td>
<td>1.57</td>
<td>0</td>
</tr>
<tr>
<td>Provides ongoing, honest and constructive feedback to student/trainees</td>
<td>58.92</td>
<td>38.53</td>
<td>2.57</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Illustrates appropriate professional and ethical conduct</td>
<td>50</td>
<td>46.42</td>
<td>3.57</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Adjusts teaching to learners' needs whilst taking account of the learning objectives/outcomes</td>
<td>46.43</td>
<td>44.84</td>
<td>7.14</td>
<td>1.79</td>
<td>0</td>
</tr>
<tr>
<td>Stimulates reflection, problem solving, self-directed, and independent learning skills</td>
<td>51.78</td>
<td>37.51</td>
<td>8.92</td>
<td>1.78</td>
<td>0</td>
</tr>
<tr>
<td>He/she is clinically competent (has sound analytical, diagnostic, therapeutic, and reasoning skills)</td>
<td>52.21</td>
<td>42.42</td>
<td>5.35</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Is a good role model for students/trainees</td>
<td>41.07</td>
<td>51.78</td>
<td>7.14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Demonstrates adequate clinical and medical knowledge</td>
<td>35.71</td>
<td>53.57</td>
<td>3.57</td>
<td>5.36</td>
<td>1.79</td>
</tr>
<tr>
<td>Remains up to date with knowledge of developments in the field</td>
<td>66.07</td>
<td>25.32</td>
<td>8.64</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Is able to teach in diverse settings (bedside, operating theatre, wards) and involves patients in teaching (if relevant)</td>
<td>32.14</td>
<td>51.79</td>
<td>14.29</td>
<td>1.79</td>
<td>0</td>
</tr>
<tr>
<td>Teaches concepts and builds skills in an organized manner</td>
<td>51.78</td>
<td>33.92</td>
<td>10.71</td>
<td>3.57</td>
<td>0</td>
</tr>
<tr>
<td>Maintains amiable, polite, and considerate attitude with students/trainees</td>
<td>33.36</td>
<td>59.44</td>
<td>7.28</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Displays appropriate and relevant sense of humor</td>
<td>19.64</td>
<td>48.21</td>
<td>19.64</td>
<td>8.93</td>
<td>3.57</td>
</tr>
</tbody>
</table>
and “pretty much of what constitutes effective clinical teaching.”

Testing of the instrument

In total, 137 specialty trainees consented to participate in the study. Of these, 83 (60.5%) were able to provide five completed assessments of the same teaching session (i.e., not assessments of five different sessions) resulting in a total data set of 415 instruments. The trainees were at different stages of their training; year 1 (10), year 2 (15) year 3 (20) year 4 (22), year 5 (9), and year 6 (7) and came from different clinical specialties: anesthesiology (9), internal medicine (16), surgery (8), gynaecology and obstetrics (13), paediatrics (18), pathology (4), radiology (5), and psychiatry (10). Trainees were assessed for a minimum duration of 15 min to a maximum of 60 min. The assessments took place on the ward, at the bedside, or while lecturing. The overall mean score across all 15 items was 8.0 (SD = 0.91) on a 10 cm visual analog scale, with 0 indicating poor and 10 excellent.

Generalizability analysis

The results of the generalizability study (Table 3) demonstrate that the proportion of the total variance due to differences in the trainees’ overall teaching performance (t) was 28%, while differences in assessors’ stringency for the same trainees (a:t) was 4%. Item variation constituted 18% of variance, while trainees’ aptitude for specific items (txi) constituted 16% of the total variance. The largest source of variance 36% resulted from assessors interpreting items differently (axi:t). The G coefficient using five assessors and 15 items is 0.92. Table 4 shows the results of the D study, with a G coefficient estimate with different numbers of assessors. Given that a reliability of 0.80 is regarded as acceptable for high stakes examinations (Richter Lagha et al. 2012), the D study shows that a minimum of two assessors is sufficient to give an acceptable reliability.

Validity

The factor analysis showed that the data decomposed into three factors that accounted for 69.72% of the total variance. The factor loadings for each item on each factor are given in Table 5. From the pattern of loadings, the three factors identified (together with the variance each accounted for and their internal consistency) were factor 1 (labeled ‘teaching and learning milieu’) variance = 42.60%; Cronbach’s α = 0.883; factor 2 (labeled ‘teaching skills’) variance = 18.90%; Cronbach’s α = 0.887; and factor 3 (labeled ‘learner oriented’) variance = 8.15%; Cronbach’s α = 0.842.

Feasibility

The minimum time taken to complete the instrument after observation of the teaching was three minutes and the maximum time was eight minutes. The mean time taken to complete the instrument was five minutes. Feedback from the
participants indicated that none of them had any difficulty in completing the instrument.

**Acceptability**

Analysis of free text indicated that both the assessors and the trainees found the instrument an acceptable method of assessment. Examples of the comments of respondents are “I think this could be really helpful for us who are trying to improve their teaching,” “for me this is a very nice and acceptable way of assessment,” “yeah it is a good way of assessment,” “acceptable and useful,” “it is a quick and easy way” and “good teaching tool.” From the content analysis, recurring words identified were acceptable, useful, and good tool. Participants did not provide any negative comments about the acceptability.

**Educational impact**

Analysis of the feedback indicated that all the participants considered it a teaching assessment which could potentially help trainees to improve their teaching skills. There was no negative feedback from the participants. Examples of the comments of the trainees were “this should become a part of our training program as it can improve our teaching,” “I want to improve my teaching and through this I can identify where I need further improvement,” “this should be ongoing to enable good teaching,” “I found it useful” and “it can definitely help in my teaching.” The recurring words identified were helpful, excellent, ongoing, and positive impact. None of the participants provided any negative feedback about the educational impact.

**Discussion**

The purpose of the Delphi study was to seek consensus of relevant experts on the attributes of effective clinical teaching, thereby addressing the issue of content validity. In the first round, the majority of the participants indicated agreement (agree/strongly agree) with the statements. However, there was disagreement about the need for humor in teaching with some participants suggesting that humor is not essential for teaching. This indicates that the Delphi study as a methodology did allow participants to differentiate between essential and desirable items. Sometimes in surveys participants can be inclined to indicate their agreement with statements because they are busy and may not have time to read all the items or perhaps it is simply easier to indicate agreement (termed as ‘acquiescence’) (Ayidiya & McClendon 1990; Locker et al. 2007). In the present study, participants indicated disagreement and undecided for some of the statements, and differentiated between statements in their levels of agreement (that is, they indicated strongly agree for some statements and agree for others) which suggests that participants were not providing acquiescent answers. In the second round, there were a few statements on which respondents indicated “undecided”. One possible reason could be that some respondents considered these characteristics as too complex for trainees, requiring further development or experience and understanding of teaching. However, the majority agreed with these statements and they were therefore retained.

The generalizability study demonstrated that the instrument provided a reliable measure of trainees’ teaching abilities as indicated by the G coefficient. Ideally trainees’ teaching ability would be expected to be the highest source of variance because it would indicate that each trainee is performing differently from other trainees. However, the highest source of variance resulted from the assessors interpreting the items differently – for example, one assessor may regard ‘providing constructive feedback’ as being more important than another examiner and therefore mark more stringently. This has been reported as the highest source of variance in another study (Copeland & Hewson 2000) but it remains unclear how this source of variance can be reduced or minimized.

In the present study, the majority of the trainees scored in the middle or high end of the scale. One plausible explanation for this is that all the trainees who volunteered for this study were actually good teachers and scored well on their teaching assessment. Another possible reason could be that trainees selected their own assessors and the familiarity of the assessor with the trainee could have influenced the scores. The latter might also account for the low variance for assessor stringency and result in a falsely high estimate of generalizability. However, the G coefficient found in this study is comparable with values reported in studies where assessors were not self-selected (Copeland & Hewson 2000; de Oliveira Filho et al. 2008). Therefore, the high level of reliability found here is not

![Table 5. Results of factor analysis.](image-url)
necessarily explained by the self-selection of the assessors by the trainees.

The purpose of the D study was to determine the G coefficient when varying the number of assessors. The results of present study showed that it is possible to conduct this teaching assessment with two assessors (as determined by D study), with an increase in the number of assessors beyond three having only a minor effect on the reliability. This instrument therefore offers a practical and feasible method of assessment in the clinical setting, particularly as it can be used for a wide range of methods of educational delivery. As we were interested in assessing the effectiveness of clinical teaching against the 15 items specific to the instrument, the effect of varying the number of items was not included in the D study.

We used exploratory factor analysis to identify the interrelationships among items on the teaching instrument to determine the construct validity. The result of factor analysis indicates that the teaching construct is multidimensional and these three factors form the underlying dimensions of the construct. Previous research (Litzeiman et al. 1998; Zuberi et al. 2007; Nation et al. 2011) which identifies the factors ‘teaching skills’ and ‘learning climate’, supports the two factors ‘teaching skills’ and ‘learning and teaching milieu’ identified in the present study, while the identification of the factor ‘learner support’ (Reese 2009) corroborates the component ‘learner-oriented’ in the present study.

Participants found it an acceptable method to use in clinical settings and although their feedback was brief, it can be argued that if participants do not like something or find it unacceptable, they will not hesitate to comment (even if it is a brief sentence or word) (Mcleod 1991).

In terms of educational impact, participants’ feedback indicates that the instrument can potentially facilitate trainees to improve their teaching performance; however, this needs to be explored further. Future research can aim to explore educational impact by comparing teaching performance of trainees over a period of time. Such an approach would require a minimum of two teaching assessments to identify and compare any change. A suitable time duration between the first and the second assessment would be needed to ensure that a true estimate of any improvement could be made; but the problem persists of determining whether any change is attributable to the use of the instrument or the result of other variables (e.g., teacher training or simply by teaching more and gaining experience).

If the content of the new instrument is compared with existing instruments, it can be observed that some of the attributes are similar (e.g., communicates effectively, provides feedback, models professional characteristics). This indicates that there is consistency in the literature regarding these attributes, further implying that these characteristics can be considered as essential for effective clinical teaching. However, existing instruments have attributes (e.g., teaches principles of cost-appropriate care, is present or has coverage for rounds, setting of expectations in the clinics such as responsibilities, assignments, and grading) which may be relevant when assessing teaching among senior doctors, but may not be essential for trainee teachers. Hence, it can be concluded that using existing instruments designed for the assessment of teaching among senior doctors cannot be directly translated to the assessment of trainee doctors, unless necessary alterations are made and the instrument is revalidated (Zibrowski et al. 2011). This supports the decision to design a new instrument specifically for the assessment of trainees.

Furthermore, this instrument can be used for assessment of clinical teaching across a wide range of clinical specialties and subspecialties; existing instruments tend to be specialty focused (Love et al. 1982; de Oliveira Filho et al. 2008; Lombarts et al. 2009) and have items which are not generalizable across specialties. In addition, this new instrument can be used by clinical teachers, peers, and students; existing instruments tend to be restricted to ratings by students only. Finally, in contrast to the existing studies, the present study examined the feasibility of using the instrument in clinical settings, demonstrated the acceptability of the instrument by both assessors and trainees, and explored the perceived educational impact of the instrument which was positive that is trainees perceived it useful in facilitating them to improve their teaching.

Limitations

The results of a Delphi study are credible if the panel of experts has relevant knowledge and experience regarding the topic (Hsu & Sandford 2007) and it involves those individuals in the decision making process who are most likely to be affected by the decision of the Delphi study (Hasson et al. 2000; Powell 2003). Furthermore a Delphi study is based on the assumption that several people are less likely to come to wrong conclusions or decisions (Hasson et al. 2000) than individuals. Whilst this study addressed both these requirements, the self-selected sample is a limitation of the present study. Nevertheless, more than 100 respondents participated in the study and those who participated in the study were directly or indirectly involved in clinical teaching and had knowledge and experience of teaching in clinical settings. In addition, respondents were from a range of clinical specialties, thereby eliminating concerns about the generalizability of the results to other specialties.

The trainees who participated in the study were from one region of one country. Furthermore, participation in the study was voluntary which could have resulted in trainees taking part who prefer to engage in teaching or who are good teachers, while omitting those who avoid clinical teaching. Future research should address these issues.

It was only possible to elicit perceived educational impact of the instrument; it was not possible to demonstrate the actual educational impact of the instrument. However, none of studies using existing instruments attempted to measure educational impact, a likely reason for this being that evaluating educational impact is complex and requires time.

Conclusion

The results provide support for the validity, reliability, feasibility, acceptability, and educational impact of this instrument. With the development of this instrument, all specialty trainees...
can be assessed in their teaching skills. The instrument can be used to provide feedback on teaching by senior doctors, peers, and students. In this way, specialty trainees can be facilitated to become better clinical teachers.

**Notes on contributors**

Dr Sona IjaZ Haider, PhD, is an Assistant Director at the Department of Medical Education and Research, Shalamar Medical and Dental College, Lahore, Pakistan.

Prof Neil Johnson, MA, BM Bch, MSc, MD, FRCP, FRCPG, FHEA, FACMed, is a Professor of Medical Education and Associate Dean, Warwick Medical School, The University of Warwick, England, UK.

Prof Jil Thistlethwaite, BSc, MBBS, PhD, MMed, FRCP, FRACP, DRCOG, FHEA is a Professor of Medical Education and Director of Centre for Medical Education Research and Scholarship at the School of Medicine, University of Queensland, Brisbane, Australia.

Dr Gay Fagan, PhD, is an Associate Professor, Assessment Co-ordinator and Lead for the MBChB written examinations, Warwick Medical School, The University of Warwick, England, UK.

Dr Muhammad Furgan Bari, MBBS, PhD, is an Assistant Professor, Department of Pathology, Dow International Medical College, Karachi, Pakistan.

**Declaration of interest:** The authors report that they have no declaration of interest.

**References**


