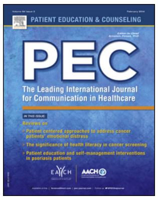
Contents lists available at ScienceDirect

Patient Education and Counseling

journal homepage: www.elsevier.com/locate/pateducou



Review

Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review[☆]

Winny Setyonugroho[a], Kieran M. Kennedy[b], Thomas J.B. Kropmans[b,*]

[a] *Faculty of Medicine and Health Sciences of the Universitas Muhammadiyah Yogyakarta, Bantul, Indonesia*
[b] *School of Medicine, College or Medicine, Nursing & Health Sciences, National University of Ireland Galway, Galway, Ireland*

ARTICLE INFO

Article history:
Received 11 September 2014
Received in revised form 27 May 2015
Accepted 6 June 2015

Keywords:
Communication skills
Systematic review
Assessment
OSCE
Undergraduate medical education

ABSTRACT

Objectives: To explore inter-rater agreement between reviewers comparing reliability and validity of checklist forms that claim to assess the communication skills of undergraduate medical students in Objective Structured Clinical Examinations (OSCEs).
Methods: Papers explaining rubrics of OSCE checklist forms were identified from Pubmed, Embase, PsycINFO, and the ProQuest Education Databases up to 2013. Included were those studies that report empirical validity or reliability values for the communication skills assessment checklists used. Excluded were those papers that did not report reliability or validity.
Results: Papers focusing on generic communication skills, history taking, physician–patient communication, interviewing, negotiating treatment, information giving, empathy and 18 other domains (ICC −0.12–1) were identified. Regarding the validity and reliability of the communication skills checklists, agreement between reviewers was 0.45.
Conclusions: Heterogeneity in the rubrics used in the assessment of communication skills and a lack of agreement between reviewers makes comparison of student competences within and across institutions difficult.
Practice implications: Consideration should be afforded to the adoption of a standardized measurement instrument to assess communication skills in undergraduate medical education. Future research will focus upon evaluating the potential impact of adoption of a standardized measurement instrument.

Contents

[☆] The content was presented at the Association for Medical Education in Europe (AMEE) 2013 conference in Prague, Czech Republic.

[*] Corresponding author. Tel.: +353 91 495478; fax: +353 91 495512.
E-mail addresses: wsetyonugroho@umy.ac.id (W. Setyonugroho), kieran.kennedy@nuigalway.ie (K.M. Kennedy), thomas.kropmans@nuigalway.ie (Thomas J.B. Kropmans).

http://dx.doi.org/10.1016/j.pec.2015.06.004
0738-3991/© 2015 Elsevier Ireland Ltd. All rights reserved.

1. Background

Physicians' communication skills (CS) have a considerable impact upon quality of health care, whereby good CS improve healthcare outcomes, such as physiologic status, pain control, and emotional health, and significantly increase patient understanding and patient satisfaction [1,2].

Effective physician–patient communication is essential in ensuring that patients adequately understand their diagnoses, treatment options, medications, plans for referral and prognosis. Dissatisfaction with physician–patient communication is known to be a leading factor influencing patients' decisions to initiate medical negligence proceedings [3,4]. Existing research demonstrates that errors in physician–patient communication include inadequate information-giving, reluctance to adopt a specific partnership style, being in a hurry and failing to respond to patients' feelings [5–7]. In clinical settings, CS take verbal, nonverbal and written forms. From the point of view of physician–patient interaction, CS can be classified according to purpose. For instance, initiation of a session, gathering information, providing structure to the interview, building a relationship, explanation and planning, closing the session and other specific issues [2,4,8,9].

Within medical educational settings, practical CS training has been shown to improve medical student performance in relationship building, time management and patient assessment [10]. According to Humphris [11] medical students' acquisition of CS is influenced not only by structured teaching sessions but also by incidental learning. The development of communication knowledge has a small, but significant, influence on performance [11].

The Objective Structured Clinical Examination (OSCE), an assessment method introduced by Harden in 1975, is the assessment tool most commonly used for assessment of clinical skills in undergraduate medical education [12]. Research suggests that the OSCE is appropriate for high-stakes assessment [13]. In addition to practical clinical skills, the OSCE can be used to assess complex CS [14]. Such assessment may take the form of OSCE stations dedicated to the assessment of CS or stations testing specific subject areas or domains of CS alongside other clinical skills. In our Medical School, the majority of OSCE stations combine both the assessment of domains of CS with assessment of a specific set of clinical skills. It is acknowledged that the combination of CS domain checklists with clinical skills checklists is likely to influence the choice and design of the assessment tools. Interpretation of student performance in such stations can be complicated by the combination of CS and clinical skills assessment, such that students may compensate between these skills to achieve a pass grade overall, whereas their performance in the individual competencies is often not immediately apparent. The OSCE itself has evolved into many variations, the Objective Structured Clinical Assessment (OSCA) and the Group Objective Structured Practical Examination (GOSPE) [15–18].

Very many different measurement instruments have been used to evaluate CS in OSCEs [19]. To examine such skills, two types of scale ratings are frequently used. The first type is that of a "behavioral checklist" and the second is a "multi-point global scale/global rating scale" (GRS). There is evidence supporting the use of global rating scales (GRS's) rather than checklists [15]. Research suggests that GRS's have higher internal consistency when compared against checklists, and furthermore, that using both GRS's and checklists in combination can improve content validity [20,21].

CS can either be assessed during real time assessments or after a recorded session. Those collecting data in real time have the potential advantage of being able to provide instant feedback to participants, whilst recorded sessions have the advantage of generating permanent data that can be used for repeated analysis [15]. Regardless of the method used, it is important for medical educators to evaluate students' CS on a number of occasions over their entire course of study so that an improvement in ability can be recognized and so that those students who are failing to progress can be identified [22]. However, heterogeneity in measurement instruments used to assess CS in OSCEs limits the comparability of student performance between examinations settings. It would be expected that most institutions would be using similar rubrics for assessment of CS in different years of their degree programmes, thus allowing easy comparison between students and of individual progress across academic years. Ideally, there would also be consistency within and between institutions in terms of the rubrics used to assess CS.

There is an existing body of research pertaining to the assessment of CS. Beck et al. [4] reviewed measurable verbal and non-verbal CS of physician–patient communication, Ong et al. [2] compared interaction analysis systems and Boon and Stewart [19] reviewed available instruments to assess physician–patient communication. Schirmer et al. [23] compared to what extend the instruments measured essential elements of communication in the family medicine context. The aim of the present review is to explore inter-rater agreement between reviewers analyzing quality and content of papers systematically by comparing whether reliability and validity of checklist forms that claim to assess the CS of undergraduate medical students in OSCEs are described appropriately in these papers. Agreement between raters about quality and content of the included papers is expressed in an intra class correlation coefficient (ICC).

2. Method

A preliminary narrative literature review, pertaining to clinical CS and OSCEs, was conducted by the Principle Investigator (PI), WS, in order to ensure that key points and conceptual framework~~s~~ were adequately covered in later search strategies. A list of keywords was developed from the results of this exercise, so that they could form the basis for a more extensive literature search detailed below.

A search was performed in order to identify studies which were published between January 1975 (first description of the OSCE) and December 2012, in peer reviewed publicly available international journals published in English. The following databases were searched: PUBMED, EMBASE, PsycINFO Ovid, and ProQuest Education Databases (consisting of ERIC, British Education Index, and the Australian Education Index).

Boolean operators (i.e. AND, OR, NOT or AND NOT) were used as conjunctions to combine or exclude keywords in a search, thereby resulting in more focused and relevant results in PUBMED. These were adapted accordingly for the other databases. The examples of search terms identified in this manner were "Objective Structure Clinical Examination", "OSCE", or any variation of OSCE including the abbreviations. This was followed by combining results, using the Boolean logic AND, with words from communication domains such as "communication", "history taking", "physician–patient relationship", "interview" or "counseling".

A series of search strategies was utilized to ensure correct results and limits were applied to remove false results. The search strategy for PUBMED is provided below. This was adapted accordingly for the other three databases.

Whilst the Boolean string operator "NOT" was applied in PUBMED, application to the other databases was problematic due to different Boolean logistics. Thus, we used reference management software, known as Zotero, to overcome this issue. The PI,

Please cite this article in press as: Setyonugroho W, et al. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. Patient Educ Couns (2015), http://dx.doi.org/10.1016/j.pec.2015.06.004

WS, carried out a manual search of the references of identified studies in order to identify further relevant studies.

We included studies which described the assessment of CS using OSCEs in undergraduate medical students. Only papers referring to undergraduate medical students were included. Studies conducted within dentistry, veterinary, pharmacy and other para-medical disciplines were excluded. Papers were included if they described OSCE stations which were entirely dedicated to the assessment of CS. Papers which described OSCE stations that assessed CS only as a component of a broader assessment, such as clinical examination or procedural skills, were also included. Studies were excluded if they did not provide empirical validity or reliability information in relation to the assessment checklist used (i.e. papers had to explicitly state the validity and/or reliability of their assessment checklist or a reference to an existing study of the validity and/or reliability of the checklists). Studies were included regardless of the nature of the specific clerkship that the OSCE was associated with and regardless of whether or not the participating students originated from the same year of study.

For the second ‘systematic review’, we included all identified CS measurement instruments (checklists) used in OSCEs. Instruments that measured CS in assessment types other than OSCEs were not included. Studies that did not provide a description of the CS measurement instrument were excluded.

Each reviewer analyzed the included literature using a data-extraction template. The template was designed using keywords and assessment rubrics found in potentially relevant papers. It consisted of 2 categories, whereby category one sampled 22 domains of CS as assessed by an examiner and category two sampled 5 domains of CS as assessed by a Standardized Patient rater (SP rater). Other information that was extracted from each paper included the study sample size (number of students), the duration of stations (recorded in minutes), the utilization or otherwise of a CS checklist and referral to any professional board or licensing bodies.

After an initial meeting to agree the meaning of each item on the data-extraction template, WS, TK, KK independently analyzed each research paper. Where two out of three reviewers were in agreement (initial agreement in percentage), these items were discussed with a view to achieving complete agreement where possible (resolved disagreement in percentage). To correct for change an Intraclass Correlation Coefficient (ICC) was calculated. Data was entered into SPSS (version 20) and the levels of agreement between reviewers for each of the 27 domains of CS were measured using ICC. Full agreement between reviewers (ICC = 1) means 100% agreement on items assessed. No agreement (ICC = 0) means that reviewers did not agree at all on the items that were assessed. An agreement of 0.45 means reviewers agreed on 45% of the items that were assessed with a correction for agreement by chance.

Ethical approval was not required for this review.

3. Results

3.1. Search results

The initial literature search identified 1998 papers (Fig. 1). After removal of duplicates, 1358 papers remained. By review of the titles and abstracts, 613 were excluded on the basis of irrelevancy. A further 557 papers were excluded as they were not related to OSCEs in undergraduate medical schools. Manual review of the titles and abstracts of the remaining papers identified a further 20 duplicates and 13 non-English language papers, all of which were excluded. In cases where it was not possible to make the decision to include or exclude a paper based upon its title and abstract alone, the full text of the paper was reviewed. Review of the full text of the remaining 87 papers revealed that 48 did not assess CS and a further 5 did not provide validity or reliability data. All of these papers were excluded, thus leaving 34 papers to be included in the review (Table 1).

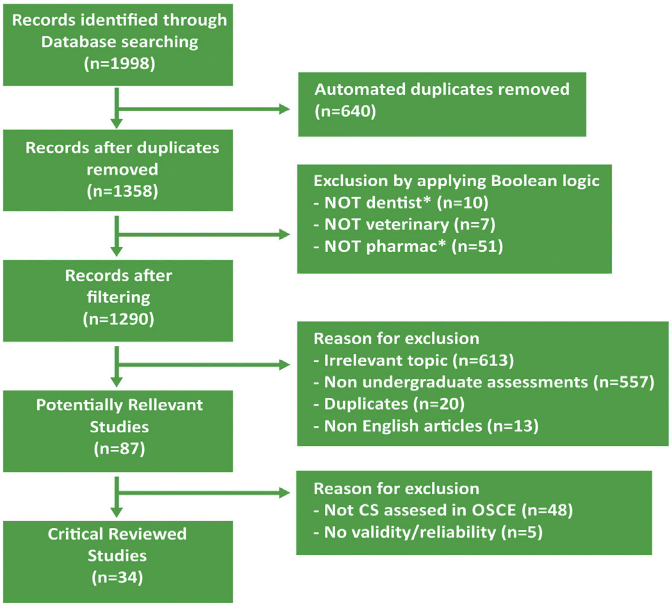

Records identified through Database searching (n=1998)
Automated duplicates removed (n=640)
Records after duplicates removed (n=1358)
Exclusion by applying Boolean logic
- NOT dentist* (n=10)
- NOT veterinary (n=7)
- NOT pharmac* (n=51)
Records after filtering (n=1290)
Reason for exclusion
- Irrelevant topic (n=613)
- Non undergraduate assessments (n=557)
- Duplicates (n=20)
- Non English articles (n=13)
Potentially Rellevant Studies (n=87)
Reason for exclusion
- Not CS assesed in OSCE (n=48)
- No validity/reliability (n=5)
Critical Reviewed Studies (n=34)

Fig. 1. The different stages of the systematic review.

3.2. Content analysis

The number of student participants in individual studies ranged from 36 to 476, with an average number of 185 students. Twenty five studies reported the duration of CS stations, with the shortest being 5 min long and the longest being 20 min long [11,24–46], whilst almost half of the reviewed papers ranged from 5 to 8 min [11,24,26–31,33,36–39,42–44]. Four studies reported short and long case scenarios with obviously different durations [11,31,32,37], in contrast to nine studies which did not report station length at all [15,47–55]. Sixteen studies reported references to validity and reliability studies of the assessment forms being used [11,15,24,26,27,30,39,40,42–46,52–54], whilst six studies reported only validity ([31,33,37,50,55], and twelve reported only reliability [25,28,32,34–36,38,41,47–49,51]. Two studies compared assessments of medical student CS across different institutions [40,52], whilst the remainder were based in a single institution. The included studies involved participants across the range of 1st year to final year. Sixteen studies report upon the assessment of 3rd year students [15,25–27,29,31,34,36–38,41,47,51–54], and six studies reported upon more than one year of study [11,15,31,37,40,54]. In two studies students were assessed only by SP raters. [50,51] In eight studies students were assessed by examiners and by SP raters [11,26,31,34,36,45,52,53]. In the remaining 24 studies students were assessed by examiners alone.

The majority of the papers examined focused upon eight domains, which included generic CS, physician–patient communication, history-taking, focused history-taking, interviewing, negotiating plan/treatment, information giving and empathy (Table 2). The papers where SP raters were involved as assessors focused mainly on generic CS and interpersonal skills. The term “generic CS” was used where reviewed papers only mentioned ‘communication skills’ without giving any additional information regarding specific descriptions of the CS domains being addressed.

Please cite this article in press as: Setyonugroho W, et al. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. Patient Educ Couns (2015), http://dx.doi.org/10.1016/j.pec.2015.06.004

Table 1
Steps conducted in the initial narrative review to retrieve the appropriate literature for the systematic critical appraisal of the literature.

1.	OSCE
2.	Objective Structured Clinical Examination
3.	OR 1–2
4.	MMI
5.	“multiple mini interview”
6.	“multiple mini-interview”
7.	OR 4–6
8.	MiniCex
9.	Mini-cex
10.	“mini Clinical Evaluation Exercise”
11.	mCEX
12.	OR 8–11
13.	OSCA
14.	“objective structured clinical assessment”
15.	OR 13–14
16.	TOSCE
17.	“team observed structured clinical encounter”
18.	OR 16–17
19.	GOSPE
20.	“group objective structured practical examination”
21.	OR 19–20
22.	3 or 7 or 12 or 15 or 18 or 21
23.	Communication
24.	“communication skills”
25.	“history taking”
26.	Consultation
27.	“consultation skills”
28.	“breaking bad news”
29.	“cross cultural”
30.	“interpersonal relation”
31.	“end of life”
32.	“informed consent”
33.	Anamnesis
34.	Interview
35.	“medical interview”
36.	“doctor-patient interaction”
37.	“doctor-patient relation”
38.	“physician–patient relation”
39.	“physician–patient interaction”
40.	Referral
41.	Counseling
42.	“non verbal communication”
43.	“electronic communication”
44.	“email communication”
45.	“doctor-nurse communication”
46.	“physician–nurse communication”
47.	“health beliefs”
48.	“treatment plan*”
49.	OR 23–48
50.	22 AND 49
51.	Dentist* (Title/Abstract)
52.	Veterinary (Title/Abstract)
53.	Pharmacy (Title/Abstract)
54.	Pharmacist (Title/Abstract)
55.	OR 51–53
56.	50 NOT 55

Regarding adherence to recognized standards, three papers reported use of the Calgary-Cambridge Observation Guide (CCOG) [37,48,49], whilst two others used the Maas-Global and revised Maas-Global (Maas-R) [33,40], and two papers used the Standardized Patient Satisfaction Questionnaire (SPSQ) [26,31]. The Patient Perception Questionnaire (PPQ), the American Board of Internal Medicine Patient Satisfaction Questionnaire (ABIM PSQ), the Liverpool Communication Skills Assessment Scale (LCAS), the Global Simulated Patient Rating Scale (GSPRS) and the WHACS mnemonic which were each used only by a single study [11,27,31]. The WHACS mnemonic provides an essential checklist for history taking on occupational and environmental health and was created by the Environmental Medicine Curriculum committee of the South Carolina Statewide Family Practice Residency Program [27,56]. Chessman [31] and Humphris [11] incorporated more than one recognized standard into their checklists.

Chessman [31], Park [34], Wong [46], Kaul [51] and McLay [53] referred to professional board/licensing bodies in describing the design of their instruments. These included the ACGME (Accreditation Council for Graduate Medical Education), NBME (National Board of Medical Examiner) and ABIM (American Board of Internal Medicine).

Fourteen papers reported the use of checklists or global rating scales [15,28,29,34,35,38,41,43,45,46,50–52,55], while global rating scales were described in 3 of the 14 papers by Park [34], Wass [43] and Hodges [15], respectively, in Park's paper students were assessed by SP raters only [34]. However we identified 11 papers not reporting any measurement instruments in terms of a documented list of items described in the paper or an appendix [24,25,30,32,36,39,42,44,47,53,54]. Based upon the information provided in the title, abstract and content, they were excluded from our search results.

3.3. Standard setting

The Maas-Global, the first available standard proven to be valid and reliable, consists of a check-list and a 20-page scoring manual, listing criteria per item [40]. The focus of this instrument is on the communication process, rather than the content, i.e. *how* questions are asked rather than *what* is asked [40]. Simone Scheffer validated a Global Rating Scale assessing empathy, degree of coherence in the interview, verbal expression and non-verbal expression [37]. In her study encounters were evaluated using the short version of the Calgary Cambridge Observation Guide. This Guide divides communication in medical settings into two broad categories: (a) interviewing the patient and (b) explanation and planning. Each of the categories has several components. For example, interviewing the patient is further divided into (a) initiating the session, (b) gathering information, (c) building relationship, and (d) explaining and planning [57]. According to the CCOG, this guide can be used as checklist for CS assessment and as feedback tool to the learner although publications on reliability and validity of the CCOG as an assessment tool are lacking. However many of the checklists we found used the CCOG as a kind of standard. The Standardized Patient Satisfaction Questionnaire (SPSQ) scores students in the following performance domains: (1) interviewing skills, (2) negotiating the diagnosis or plan, (3) gathering case-specific content information, (4) responding to the patient's emotions, and (5) student's overall performance. Pearson Product-Moment correlations were calculated for each of these domains [26].

3.4. Reviewer agreement

Agreement between our reviewers, expressed in an Intraclass Correlation Coefficient (ICC) on the CS domains, ranged from −0.12 to 1 and the ICC on all CS domains was 0.81, while total ICC on all marked items was 0.68 (Table 3).

Agreement improved after the reviewers discussed items whereby only two out of three agreed initially. For the purposes of presenting the results, the situation where reviewers were in full agreement prior to such discussion is termed ‘‘initial agreement’’. The situation whereby reviewers achieved full agreement after discussing the disagreed item(s) is termed ‘‘resolved disagreement’’. The comparison between initial agreement (17%) and resolved disagreement (83%) for measurement instruments amongst reviewers is illustrated in Fig. 2. For CS domains, initial agreement was 33% and this increased to 67% upon discussion. Meanwhile ‘n of student’ and ‘duration of station’ had low percentage of resolved disagreement and ‘validity/reliability’ were 50%.

Please cite this article in press as: Setyonugroho W, et al. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. Patient Educ Couns (2015), http://dx.doi.org/10.1016/j.pec.2015.06.004

Table 2
Details of papers included in the systematic review and an overview of the communication skills domains reported in each.

Author, year	*n* Of students	Length of stations (min)	Validity, Reliability (V = validity, R = reliability)	Measurement Instruments	Study Year	Examiners domains	SP raters domains	Professional boards or organizations
Al-Naami, 2008	64	5	V, R	n/a	Final year surgical clerkship	Generic CS, history taking, focused history taking		
Bergus et al., 2009	51	15	R	n/a	3rd	Generic CS, Physician–patient communication		
Blue et al., 1998	89	n/a	R	n/a	3rd	History taking, focused history taking		
Blue et al., 2000	476	8	V, R	SPSQ	3rd	Focused history taking, interview, negotiating plan/treatment, information giving, empathy, emotion/ respond of emotion	Generic CS, interpersonal skills	
Blue et al., 2000	205	8	V, R	WHACS, checklist	3rd	Focused history taking, interview, empathy		
Boehlecke et al., 1996	155	5	R	Checklist	2nd	Focused history taking		
Bosse et al., 2012	103	n/a	R	Calgary-Cambridge	5th	Physician–patient communication, history taking, counseling, consultation, health beliefs, interpersonal skills		
Cave et al., 2007	396	5	V	Checklists	3rd	Generic CS, introduction, history taking, focused history taking, negotiating plan/treatment, information giving		
Chesser et al., 2004	192	5	V, R	n/a	Penultimate undergraduate year	Generic CS, history taking, focused history taking		
Chessman et al., 2003	127	8 and 15	V	SPSQ, PPQ, ABIM PSQ	3rd, 4th	Generic CS	Generic CS, interpersonal skills	ABIM
Harasym et al., 2008	190	n/a	R	Calgary-Cambridge	Family Medicine rotation	Focused history taking		
Ho et al., 2010	57	n/a	V	Checklists	5th		Generic CS	
Hodges and McIlroy, 2003	57	10	V, R	Checklist, global rating scale	3rd, 4th	Non-verbal communication, empathy		
Huang et al., 2010	256	10 and 20	R	n/a	7th	Generic CS, history taking,		
Humphris, 2002	383	5 and 10	V, R	LCAS, GSPRS	1st, 2nd	Introduction, non-verbal communicationunicating, empathy	Generic CS	
Jacobs et al., 2004	356	5	V	Maas-R	5th	Introduction, history taking, focused history taking, negotiating plan/ treatment, information giving, interpersonal skills, empathy		
Kaul et al., 2012	279	n/a	R	Checklists	3rd		Generic CS, history taking	ACGME
Mazor et al., 2005	100	n/a	V, R	Checklists	3rd	Negotiating plan/treatment, information giving, health beliefs, empathy	Generic CS, interpersonal skills, health beliefs, empathy	
McLay et al., 2002	82	n/a	V, R	n/a	3rd	Interview	Generic CS, interpersonal skills	NBME
Park et al., 2004	286	15	R	Checklists, global rating scale	3rd	History taking, focused history taking, interview	Generic CS, interpersonal skills	NBME
Regehr et al., 1999	161	n/a	V, R	n/a	2nd, 3rd	Generic CS, history taking		
Robins et al., 2001	71	20	R	Checklists	4th	Cross-cultural communication, health beliefs	Health beliefs	
Rosebraugh et al., 1997	196	8	R	n/a	3rd	History taking		

Table 2 (*Continued*)

Author, year	*n* Of students	Length of stations (min)	Validity, Reliability (V = validity, R = reliability)	Measurement Instruments	Study Year	Examiners domains	SP raters domains	Professional boards or organizations
Scheffer et al., 2008	113	5 and 8	V	Calgary-Cambridge	2nd, 3rd	Generic CS, interview, Non-verbal communication, empathy, micro expression		
Thistlethwaite, 2002	194	6	R	Checklists	3rd	History taking, negotiating plan/ treatment, information giving		
Troncon, 2006	36	7	V, R	n/a	4th	Physician–patient communication, history taking		
Van Dalen et al., 2002	161	15	V, R	Maas-global	4th, 6th	Generic CS, history taking, focused history taking, negotiating plan/ treatment, information giving, breaking bad news		
Verma and Singh, 1994	40	n/a	V	Checklists	Final year	Generic CS, Information giving		
Volkan et al., 2004	169	20	R	Checklists	3rd	Physician–patient communication, history taking		
Walters et al., 2005	128	6	V, R	n/a	4th	Generic CS, history taking, phone/ electronic communication		
Wass and et al., 2001	214	7	V, R	n/a	Final MBBS examination	Generic CS		
Wass and Jolly, 2001	155	8	V, R	Global rating scale	Final MBBS examination	Generic CS		
Wilkerson et al., 2010	322	15	V, R	Checklists	Senior medical students	Generic CS, negotiating plan/treatment, counseling, health beliefs, empathy	Interpersonal skills, empathy, health beliefs	
Wong et al., 2007	439	10	V, R	Checklists	Final year	Physician–patient communication, information giving, taking consent, breaking bad news, advising/handle family, interpersonal skills		ACGME

Abbreviations list: SP: standardized patient; LCAS: Liverpool communication skills assessment scale; Generic CS: generic communication skills; GSPRS: Global simulated patient rating scale; SPSQ: standardized patient satisfaction questionnaire; ABIM: American board of internal medicine; PPQ: patient perception questionnaire; ACGME: accreditation council for graduate medical education; PSQ: patient satisfaction questionnaire; NBME: National board of medical examiners; WHACS: a mnemonic, provide a few essential questions on occupational and environmental exposures.

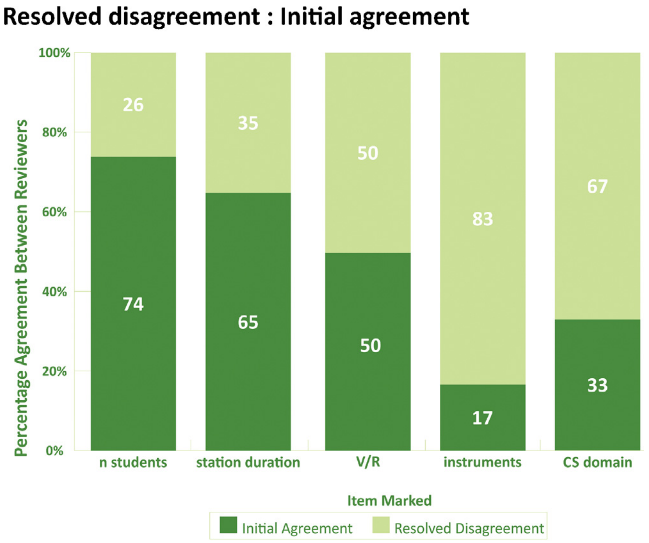


Fig. 2. Comparison between resolved disagreement and initial agreement amongst reviewers. * V/R = validity/reliability, CS domains = communication skills domains.

4. Discussion and conclusion

4.1. Discussion

The most striking finding of our study is a demonstrated absence of consensus in rubrics used to assess CS in undergraduate medical education worldwide. Furthermore, it is apparent that there is a clear absence of consensus between researchers in medical education in their interpretation of terminology and in their determination of performance standards in the assessment of CS in different settings. The OSCE is widely utilized to assess CS at undergraduate and postgraduate levels [55]. It is likely that significant heterogeneity exists in teaching and assessment of CS across different institutions as well as across different years of the curriculum within the same institution. Tone possible explanation is that those who use established local instruments do not frequently publish their adaptation or validation in their circumstances. It is perhaps not surprising that there are differences between tools designed to assess different CS. Similarly, it is not surprising that there are differences in tools designed to assess measurement of blood pressure and those used to assess performance of basic life support. However, the demonstrated lack of agreement and transparency in the use of terminology and lack of published reliability and validity of any type of OSCE station is of concern. This absence of standardization of assessment rubrics in undergraduate medical education precludes the comparison of outcomes across assessment settings. This study highlights the absence of an agreed gold standard for the assessment of CS of undergraduate medical students. This finding is of particular note in the context of the existence of the first reliable and valid measurement tool, known as the MAAS Global [58].

We identified only 9 papers (27%) which did not report station duration and 4 studies (11%) which reported two different durations (a short case and a long case approach). These results align with the findings of Patricio et al. [59] who concluded that only 30% of papers reported station duration. This finding is of concern because station duration is known to be one determinant of station reliability. Effective assessment of CS requires enough time to adequately cover the objective of the communication. Thus, we recommend that future research pertaining to the assessment of CS should always report station duration and the objective of the communication so that research findings can be more easily compared and synthesized.

The majority of included studies did not clearly report measurement instruments and the underlying construct of various CS domains was unclear in 19 out of 23 (83%) papers (Fig. 2). It is apparent from Fig. 2 that the included research on CS assessment can very easily be misinterpreted, even by expert reviewers. For instance, reviewer disagreement upon the number of student participants, an item that should be very clear, was only resolved after two meetings.

Difficulty arose during the reviewer analysis of the papers with respect to the CS domains that were measured and the interpretation of the terminology used to describe such domains. The ICC for all domains described in the papers included in the review was 0.81. With respect to the domain of 'focused interview,' we didn't achieve agreement (ICC −0.12). This finding is explained by the absence of any description of the 'focused interview' domain in all of those papers [26,27,33,34,45,49]. Full agreement was reached in only three of the included papers. Each of these papers described only one domain of CS. The crucial omission of clear descriptions of CS domains was previously described by Boon and Stewart [19], Beck et al. [4] and Cegala and Lenzmeier Broz [60,61]. A clear description of the object or underlying concepts in the relationship with empirical indicators is the single most important requirement for assessment [62–64]. If the concept that is to be assessed is not clearly defined and clear indicators are not included, then it cannot be adequately measured. We suggest that educational decisions drawn from flawed measures are unreliable.

The majority of included studies focused upon physician–patient interaction. However, in reality, physicians must also be able to communicate effectively with other physicians, with nurses and with other stakeholders [65–70]. Our review demonstrates a notable paucity of published research in this field. Furthermore, we identified only one study which explored the assessment of the use of phone/electronic communication at undergraduate level [42]. Other forms of communication include interpersonal skills, non-verbal communication, micro expression and empathy. We identified eight papers which studied the assessment of empathy, suggesting that this domain of CS is an apparent priority for researchers [11,15,26,27,33,37,45,52]).

Reliability of results and validity of CS are essential to the assessment of student competence [64]. There are other opinions according to which reliability of results is a prerequisite of validity while others mention that reliability of results is necessary but not sufficient for the sole support of validity [71]. We found that 16 (47%) of the reviewed papers reported both reliability of results and validity (internal consistency), whilst the remainder reported only one of these two measures. It was notable that the majority of papers did not refer to a recognized Gold standard with a view to improving the construct validity of each assessment form used.

In contrast, the University of Maastricht developed a unique and validated instrument, currently known as the MAAS-Global, which was first reported back in 1990 [58,72]. This instrument is being used, in real-time and in recorded sessions, to assess students, physicians and/or nurses at only a small number of institutions in different countries. However, it is important to recognize that it is not being more widely adopted as a gold standard [73–75]. It is apparent from this review that the majority of CS assessment is based upon the individual development of unique measurement instruments which are used only at local level. On the contrary, we only found three instruments the Maas-Global (including Maas-R), the SPSQ, and the CCOG which were reported in more than one study. The CCOG is actually a guideline which was not designed to be used as a validated assessment instrument. As demonstrated in previous reviews published in 1998 and 2002, the present review again identified a failure to adopt existing validated instruments [4,19]. Inability to reproduce

Table 3
Communication skills domains agreement between 3 reviewers ICC = intra class correlation coefficient Full agreement between reviewers (ICC = 1) means 100% agreement on items assessed. No agreement (ICC = 0) reviewers don't agree at all on items assessed. An agreement of 0.45 means reviewers agreed on 45% f the items assessed with a correction for agreement by chance (ICC).

	Examiners domains												
	Generic CS	Doctor-patient communication	Introduction	History taking	Focused history taking	Interview	Focused interview	Negotiating plan/Treatment	Taking consent	Information giving	Counseling	Consultation	Breaking bad news
Percentage of papers	52	34	16	58	41	25	9	30	9	27	14	10	11
ICC	0.83	0.69	0.86	0.81	0.90	0.82	−0.12	0.85	0.73	0.92	0.74	0.66	0.90

	SP raters domain													
	Cross cultural communication	Health beliefs	Advising/ handle family	Interpersonal skills	Non-verbal communication	Empathy	Micro expression	Emotion/ respond emotion	Phone/electronic communication	Generic CS	History taking	Interpersonal skills	Empathy	Health beliefs
Percentage of papers	7	20	6	17	17	30	6	8	7	31	6	25	14	15
ICC	0.90	0.85	1	0.82	0.82	0.93	1	0.82	0.90	0.91	1	0.92	0.75	0.89

results across assessments precludes meaningful interpretation of results [76].

Whilst some time has passed since the aforementioned reviews in 1998 and 2002 were carried out, the two main problems of how to define an appropriate learning outcome of a specific CS domain and an appropriate method of measurement still exist in 2015. Rather than repeatedly creating new assessment forms, researchers and educators need to work together in order to agree upon the definitions of learning outcomes and CS domains, so that gold standard CS measurement instruments can be developed. Bloch and Norman [76] doubt whether competence can be measured with a single scale (i.e. one measurement instrument for all CS), as opposed to a unique scale (i.e. a specific measurement for each specific domain of CS) for different specialties and different practice conditions. While 're-inventing the wheel' is not necessary, any effort to incorporate or modify existing instruments in order to fit into different specialties and practice conditions will be valuable for future development of undergraduate medical education. The Step2CS is a high stakes CS assessment tool. The Clinical Skills Review (CSR) is an Interactive Internet based preparatory site for the United States Medical Licensing Examination (USMLE) Step 2 Clinical Skills (CS) live exam. CSR offers a specialized learning environment that is aligned with the rules and regulations set forth by the official exam provider. We agree that there should be alignment between undergraduate CS training with the expected learning outcomes and assessment goals of these high stake licensing exams. In short, we suggest that standardization (i.e. uniform use of valid, reliable and aligned CS measurement instruments) and alignment of undergraduate and postgraduate communication skills training is necessary in order to sufficiently meet the requirements of professional practice. Whilst global standardization might be very challenging, we wish to highlight the importance of standardization and appropriate use of statistics as a prerequisite for student outcome comparison between and across local and national settings [77].

Limitations of this systematic review include the exclusion of studies published in languages other than English and those not pertaining to undergraduate medical students, thus it may not be appropriate to generalize results to assessment in other student populations and settings. In retrospect, we did not adequately take into consideration the importance of aligning postgraduate and undergraduate training and assessment of CS and the use of frameworks like CanMeds and others to highlight appropriate 'top down' alignment of CS training. Furthermore, despite rigorous research methods, incomplete retrieval of published literature is possible. Despite CS being an important competence for students to master, the assessment of CS continues to be a challenging endeavor. In the US it is now mandated that medical students and residents have CS training and the variety and variances in this training are as numerous as there are programs. Internationally, there are still medical schools that have not incorporated CS training in a formal way. Clearly there is work to be done and reviewing what we know to work well is important, starting with a clear description of the underlying concepts.

4.2. Conclusion

We demonstrate a clear absence of consensus between researchers in their interpretation and definition of domains of CS. Included papers generally failed to satisfactorily identify the underlying constructs and learning outcomes that were being assessed. Terminology was not uniformly employed across included papers.

Furthermore, there was poor consistency with respect to the use of Likert scales and global ratings scales, despite this issue having been previously identified [19]. A valid and reliable measurement instrument, such as the Maas-Global (http://bit.ly/1xQXAnS), is not universally accepted and this paper promotes calibration of communication skills using this valid and reliable standard.

4.3. Practice implications

Future research should focus upon the comparison of the clinical skills stations in our and other institutions using the Maas-Global as a standard to calibrate existing CS items in each of the assessment forms, so that measures of CS become interchangeable and comparable within and between institutions. We suggest that such calibration could be based upon the Maas-Global.

Conflict of interest statement

All three authors declare no conflict of interest.

Notes on contributors

Winny Setyonugroho, is Lecturer in Health Informatics, Department Health Informatics, Faculty of Medicine & Health Sciences, Universitas Muhammadiyah Yogyakarta, Indonesia and also a PhD student, domain Medical Informatics & Medical Education, Department of Medicine, School of Medicine, College of Medicine, Nursing & Health Sciences National University of Ireland Galway (supervised by Dr Thomas JB Kropmans).

Kieran Kennedy is Lecturer in Clinical Methods and Clinical Practice, Department of Medicine, School of Medicine, College of Medicine, Nursing & Health Sciences National University of Ireland Galway. He holds specialist training and experience in the field of communication skills education and is also a practicing General Practitioner.

Thomas Kropmans is Senior Lecturer Medical Informatics & Medical Education, Domain of Medical Informatics & Medical Education, Department of Medicine, School of Medicine, College of Medicine, Nursing & Health Sciences National University of Ireland Galway.

Acknowledgement

The lead author, Winny Setyonugroho, received a PhD scholarship from the Directorate General of Higher Education, Ministry of National Education and Culture of Indonesia.

References

References for the study

[1] Stewart MA. Effective physician–patient communication and health outcomes: a review. Can. Med. Assoc. J. 1995;152(9):1423–33.
[2] Ong LM, de Haes JC, Hoos AM, Lammes FB. Doctor-patient communication: a review of the literature. Soc. Sci. Med. 1995;40(7):903–18.
[3] Phillips C. Communication: the first tool in risk management for long-term care. J. Am. Med. Dir. Assoc. 2004;5(2):123–6.
[4] Beck RS, Daughtridge R, Sloane PD. Physician–patient communication in the primary care office: a systematic review. J. Am. Board Fam. Pract. 2002;15(1):25–38.
[5] Levinson W. Physician–patient communication. A key to malpractice prevention. J. Am. Med. Assoc. 1994;272(20):1619–20.
[6] Maguire P, Pitceathly C. Key communication skills and how to acquire them. Br. Med. J. 2002;325(7366):697–700.
[7] Maguire P, Pitceathly C. Managing the difficult consultation. Clin. Med. 2003;3(6):532–7.
[8] Noble LM, Kubacki A, Martin J, Lloyd M. The effect of professional skills training on patient-centredness and confidence in communicating with patients. Med. Educ. 2007;41(5):432–40.
[9] Bowleg L, Valera P, Teti M, Tschann JM. Silences, gestures, and words: nonverbal and verbal communication about HIV/AIDS and condom use in black heterosexual relationships. Health Commun. 2010;25(1):80–90.
[10] Yedidia MJ, Gillespie CC, Kachur E, Schwartz MD, Ockene J, Chepaitis AE, et al. Effect of communications training on medical student performance. J. Am. Med. Assoc. 2003;290(9):1157–65.
[11] Humphris GM. Communication skills knowledge, understanding and OSCE performance in medical trainees: a multivariate prospective study using structural equation modelling. Med. Educ. 2002;36(9):842–52.
[12] Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br. Med. J. 1975;1(5955):447–51.
[13] Turner JL, Dankoski ME. Objective structured clinical exams: a critical review. Fam. Med. 2008;40(8):574–8.
[14] Hodges B, Turnbull J, Cohen R, Bienenstock A, Norman G. Evaluating communication skills in the OSCE format: reliability and generalizability. Med. Educ. 1996;30(1):38–43.
[15] Hodges B, McIlroy JH. Analytic global OSCE ratings are sensitive to level of training. Med. Educ. 2003;37(11):1012–6.
[16] Kogan JR, Bellini LM, Shea JA. Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship. Acad. Med. 2003;78(10 Suppl):S33–5.
[17] Khattab AD, Rawlings B. Assessing nurse practitioner students using a modified objective structured clinical examination (OSCE). Nurse Educ. Today 2001;21(7):541–50.
[18] Khattab AD, Rawlings B. Use of a modified OSCE to assess nurse practitioner students. Br. J. Nurs. 2008;17(12):754–9.
[19] Boon H, Stewart M. Patient-physician communication assessment instruments: 1986 to 1996 in review. Patient Educ. Couns. 1998;35(3):161–76.
[20] Barry M, Bradshaw C, Noonan M. Improving the content and face validity of OSCE assessment marking criteria on an undergraduate midwifery programme: a quality initiative. Nurse Educ. Pract. 2013;13(5):477–80.
[21] Moineau G, Power B, Pion AM, Wood TJ, Humphrey-Murto S. Comparison of student examiner to faculty examiner scoring and feedback in an OSCE. Med. Educ. 2011;45(2):183–91.
[22] Chang A, Boscardin C, Chou CL, Loeser H, Hauer KE. Predicting failing performance on a standardized patient clinical performance examination: the importance of communication and professionalism skills deficits. Acad. Med. 2009;84(10 Suppl.):S101–4.
[23] Schirmer JM, Mauksch L, Lang F, Marvel MK, Zoppi K, Epstein RM, et al. Assessing communication competence: a review of current tools. Fam. Med. 2005;37(3):184–92.
[57] Amin Z, Eng KH. Basics in Medical Education, World Scientific. In: Appendix A: Calgary-Cambridge Observation Guide; 2002. Available from: http://www.worldscientific.com/doi/pdf/10.1142/9789812795472_bmatter.
[58] van Es JM, Schrijver CJ, Oberink RH, Visser MR. Two-dimensional structure of the MAAS-Global rating list for consultation skills of doctors. Med. Teach. 2012;34(12):e794–9.
[59] Patricio MF, Juliao M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? Med. Teach. 2013;35(6):503–14.
[60] Cegala DJ, Gade C, Lenzmeier Broz S, McClure L. Physicians' and patients' perceptions of patients' communication competence in a primary care medical interview. Health Commun. 2004;16(3):289–304.
[61] Cegala DJ, Lenzmeier Broz S. Physician communication skills training: a review of theoretical backgrounds, objectives and skills. Med. Educ. 2002;36(11):1004–16.
[62] Beckman TJ, Cook DA. Educational epidemiology. J. Am. Med. Assoc. 2004;292(24):2969 (author reply 70-1).
[63] 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.
[64] Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? A review of the published instruments. J. Gen. Intern. Med. 2004;19(9):971–7.
[65] Fallowfield L, Jenkins V. Effective communication skills are the key to good cancer care. Eur. J. Cancer 1999;35(11):1592–7.
[66] Fallowfield L, Jenkins V. Acronymic trials: the good, the bad, and the coercive. Lancet 2002;360(9346):1622.
[67] Fallowfield L, Jenkins V. Communicating sad, bad, and difficult news in medicine. Lancet 2004;363(9405):312–9.
[68] Fallowfield L, Jenkins V. Current concepts of communication skills training in oncology. Recent Results Cancer Res. 2006;168:105–12.
[69] Vazirani S, Hays RD, Shapiro MF, Cowan M. Effect of a multidisciplinary intervention on communication and collaboration among physicians and nurses. Am. J. Crit. Care 2005;14(1):71–7.
[70] Beuscart-Zephir MC, Pelayo S, Anceaux F, Meaux JJ, Degroisse M, Degoulet P. Impact of CPOE on doctor-nurse cooperation for the medication ordering and administration process. Int. J. Med. Inform. 2005;74(7–8):629–41.
[71] Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am. J. Med. 2006;119(2):166.e7–166.e16.
[72] Kraan HF, Crijnen AA, de Vries MW, Zuidweg J, Imbos T, Van der Vleuten CP. To what extent are medical interviewing skills teachable? Med. Teach. 1990;12(3–4):315–28.
[73] Essers G, Kramer A, Andriesse B, van Weel C, van der Vleuten C, van Dulmen S. Context factors in general practitioner-patient encounters and their impact on assessing communication skills—an exploratory study. BMC Fam. Pract. 2013;14:65.

[74] Hobma S, Ram P, Muijtjens A, van der Vleuten C, Grol R. Effective improvement of doctor-patient communication: a randomised controlled trial. Br. J. Gen. Pract. 2006;56(529):580–6.

[75] Reinders ME, Blankenstein AH, van Marwijk HW, Knol DL, Ram P, van der Horst HE, et al. Reliability of consultation skills assessments using standardised versus real patients. Med. Educ. 2011;45(6):578–84.

[76] Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. Med. Teach. 2012;34(11):960–92.

[77] step2cs.net Copyright© 2009–2015 Clinical Skills Review, LLC. All Rights Reserved. Legal@Step2CS.net.

References used for the review

[24] Al-Naami MY. Reliability, validity, and feasibility of the Objective Structured Clinical Examination in assessing clinical skills of final year surgical clerkship. Saudi Med. J. 2008;29(12):1802–7.

[25] Bergus GR, Woodhead JC, Kreiter CD. Trained lay observers can reliably assess medical students' communication skills. Med. Educ. 2009;43(7):688–94.

[26] Blue AV, Chessman AW, Gilbert GE, Mainous 3rd AG. Responding to patients' emotions: important for standardized patient satisfaction. Fam. Med. 2000;32(5):326–30.

[27] Blue AV, Chessman AW, Gilbert GE, Schuman SH, Mainous AG. Medical students' abilities to take an occupational history: use of the WHACS mnemonic. J. Occup. Environ. Med. 2000;42(11):1050–3.

[28] Boehlecke B, Sperber AD, Kowlowitz V, Becker M, Contreras A, McGaghie WC. Smoking history-taking skills: a simple guide to teach medical students. Med. Educ. 1996;30(4):283–9.

[29] Cave J, Washer P, Sampson P, Griffin M, Noble L. Explicitly linking teaching and assessment of communication skills. Med. Teach. 2007;29(4):317–22.

[30] Chesser AM, Laing MR, Miedzybrodzka ZH, Brittenden J, Heys SD. Factor analysis can be a useful standard setting tool in a high stakes OSCE assessment. Med. Educ. 2004;38(8):825–31.

[31] Chessman AW, Blue AV, Gilbert GE, Carey M, Mainous 3rd AG. Assessing students' communication and interpersonal skills across evaluation settings. Fam. Med. 2003;35(9):643–8.

[32] Huang CC, Chan CY, Wu CL, Chen YL, Yang HW, Chen CH, et al. Assessment of clinical competence of medical students using the objective structured clinical examination: first 2 years' experience in Taipei Veterans General Hospital. J. Chin. Med. Assoc. 2010;73(11):589–95.

[33] Jacobs JC, Denessen E, Postma CT. The structure of medical competence and results of an OSCE. Neth. J. Med. 2004;62(10):397–403.

[34] Park RS, Chibnall JT, Blaskiewicz RJ, Furman GE, Powell JK, Mohr CJ. Construct validity of an objective structured clinical examination (OSCE) in psychiatry: associations with the clinical skills examination and other indicators. Acad. Psychiatry 2004;28(2):122–8.

[35] Robins LS, White CB, Alexander GL, Gruppen LD, Grum CM. Assessing medical students' awareness of and sensitivity to diverse health beliefs using a standardized patient station. Acad. Med. 2001;76(1):76–80.

[36] Rosebraugh CJ, Speer AJ, Solomon DJ, Szauter KE, Ainsworth MA, Holden MD, et al. Setting standards and defining quality of performance in the validation of a standardized-patient examination format. Acad. Med. 1997;72(11):1012–4.

[37] Scheffer S, Muehlinghaus I, Froehmel A, Ortwein H. Assessing students' communication skills: validation of a global rating. Adv. 146?Health Sci. Educ. Theory Pract. 2008;13(5):583–92.

[38] Thistlethwaite JE. Developing an OSCE station to assess the ability of medical students to share information and decisions with patients: issues relating to interrater reliability and the use of simulated patients. Educ. Health (Abingdon) 2002;15(2):170–9.

[39] Troncon LE. Significance of experts' overall ratings for medical student competence in relation to history-taking. Sao Paulo Med. J. 2006;124(2):101–4.

[40] van Dalen J, Kerkhofs E, van Knippenberg-Van Den Berg BW, van Den Hout HA, Scherpbier AJ, van der Vleuten CP. Longitudinal and concentrated communication skills programmes: two dutch medical schools compared. Adv. Health Sci. Educ. Theory Pract. 2002;7(1):29–40.

[41] Volkan K, Simon SR, Baker H, Todres ID. Psychometric structure of a comprehensive objective structured clinical examination: a factor analytic approach. Adv. Health Sci. Educ. Theory Pract. 2004;9(2):83–92.

[42] Walters K, Osborn D, Raven P. The development, validity and reliability of a multimodality objective structured clinical examination in psychiatry. Med. Educ. 2005;39(3):292–8.

[43] Wass V, Jolly B. Does observation add to the validity of the long case? Med. Educ. 2001;35(8):729–34.

[44] Wass V, McGibbon D, Van der Vleuten C. Composite undergraduate clinical examinations: how should the components be combined to maximize reliability? Med. Educ. 2001;35(4):326–30.

[45] Wilkerson L, Fung CC, May W, Elliott D. Assessing patient-centered care: one approach to health disparities education. J. Gen. Intern. Med. 2010;25(Suppl. 2):S86–90.

[46] Wong ML, Fones CS, Aw M, Tan CH, Low PS, Amin Z, et al. Should non-expert clinician examiners be used in objective structured assessment of communication skills among final year medical undergraduates? Med. Teach. 2007;29(9):927–32.

[47] Blue AV, Stratton TD, Plymale M, DeGnore LT, Schwartz RW, Sloan DA. The effectiveness of the structured clinical instruction module. Am. J. Surg. 1998;176(1):67–70.

[48] Bosse HM, Schultz JH, Nickel M, Lutz T, Moltner A, Junger J, et al. The effect of using standardized patients or peer role play on ratings of undergraduate communication training: a randomized controlled trial. Patient Educ. Couns. 2012;87(3):300–6.

[49] Harasym PH, Woloschuk W, Cunning L. Undesired variance due to examiner stringency/leniency effect in communication skill scores assessed in OSCEs. Adv. Health Sci. Educ. Theory Pract. 2008;13(5):617–32.

[50] Ho MJ, Yao G, Lee KL, Hwang TJ, Beach MC. Long-term effectiveness of patient-centered training in cultural competence: what is retained? What is lost? Acad. Med. 2010;85(4):660–4.

[51] Kaul P, Barley G, Guiton G. Medical student performance on an adolescent medicine examination. J. Adolesc. Health 2012;51(3):299–301.

[52] Mazor KM, Ockene JK, Rogers HJ, Carlin MM, Quirk ME. The relationship between checklist scores on a communication OSCE and analogue patients' perceptions of communication. Adv. Health Sci. Educ. Theory Pract. 2005;10(1):37–51.

[53] McLay RN, Rodenhauser P, Anderson DS, Stanton ML, Markert RJ. Simulating a full-length psychiatric interview with a complex patient: an OSCE for medical students. Acad. Psychiatry 2002;26(3):162–7.

[54] Regehr G, Freeman R, Hodges B, Russell L. Assessing the generalizability of OSCE measures across content domains. Acad. Med. 1999;74(12):1320–2.

[55] Verma M, Singh T. Communication skills in clinical practice fad or necessity? Indian Pediatr. 1994;31(2):237–8.

[56] Schuman SH, Simpson Jr WM. WHACS your patients. J. Occup. Environ. Med. 1999;41(10):829.