

RESEARCH ARTICLE

# Assessing Communication Skills of Medical Students in Objective Structured Clinical Examinations (OSCE) - A Systematic Review of Rating Scales

Musa Cömert, Jördis Maria Zill, Eva Christalle, Jörg Dirmaier, Martin Härter, Isabelle Scholl\*

Department of Medical Psychology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

\* [i.scholl@uke.de](mailto:i.scholl@uke.de)



CrossMark  
click for updates

OPEN ACCESS

**Citation:** Cömert M, Zill JM, Christalle E, Dirmaier J, Härter M, Scholl I (2016) Assessing Communication Skills of Medical Students in Objective Structured Clinical Examinations (OSCE) - A Systematic Review of Rating Scales. PLoS ONE 11(3): e0152717. doi:10.1371/journal.pone.0152717

**Editor:** Robert K Hills, Cardiff University, UNITED KINGDOM

**Received:** October 16, 2015

**Accepted:** March 17, 2016

**Published:** March 31, 2016

**Copyright:** © 2016 Cömert et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This project was funded by the German Federal Ministry of Education and Research (grant number 01GX1043), [www.bmbf.de](http://www.bmbf.de). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** MC and EC declare that they have no competing interests. MH is co-PI in a research project funded by Mundipharma GmbH, a pharmaceutical company; IS conducted one

## Abstract

### Background

Teaching and assessment of communication skills have become essential in medical education. The Objective Structured Clinical Examination (OSCE) has been found as an appropriate means to assess communication skills within medical education. Studies have demonstrated the importance of a valid assessment of medical students' communication skills. Yet, the validity of the performance scores depends fundamentally on the quality of the rating scales used in an OSCE. Thus, this systematic review aimed at providing an overview of existing rating scales, describing their underlying definition of communication skills, determining the methodological quality of psychometric studies and the quality of psychometric properties of the identified rating scales.

### Methods

We conducted a systematic review to identify psychometrically tested rating scales, which have been applied in OSCE settings to assess communication skills of medical students. Our search strategy comprised three databases (EMBASE, PsycINFO, and PubMed), reference tracking and consultation of experts. We included studies that reported psychometric properties of communication skills assessment rating scales used in OSCEs by examiners only. The methodological quality of included studies was assessed using the COSMIN based Standards for the selection of health status Measurement Instruments (COSMIN) checklist. The quality of psychometric properties was evaluated using the quality criteria of Terwee and colleagues.

### Results

Data of twelve studies reporting on eight rating scales on communication skills assessment in OSCEs were included. Five of eight rating scales were explicitly developed based on a

physician training in shared-decision making within this research project. MH is PI in a research project funded by Lilly Pharma, a pharmaceutical company; JD and JZ currently work on this research project. The authors did not receive funding from Mundipharma GmbH nor from Lilly Pharma for this systematic review. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

specific definition of communication skills. The methodological quality of studies was mainly poor. The psychometric quality of the eight rating scales was mainly intermediate.

## Discussion

Our results reveal that future psychometric evaluation studies focusing on improving the methodological quality are needed in order to yield psychometrically sound results of the OSCEs assessing communication skills. This is especially important given that most OSCE rating scales are used for summative assessment, and thus have an impact on medical students' academic success.

## Introduction

In the 21st century, teaching and assessment of communication skills in medical schools are well recognized [1]. Effective communication is considered to be one of the most important skills of a physician [2]. According to the Accreditation Council for Graduate Medical Education (ACGME), the American Board of Medical Specialties (ABMS), the Association of American Medical Colleges (AAMC), the General Medical Council (GMC), and the World Federation for Medical Education (WFME) communication and interpersonal skills are among the essential competencies to be taught in medical and residency programs [3–7]. Over the years, several international consensus statements have been published, which aim to provide educators with knowledge in development, implementation and evaluation of communication-oriented medical curricula [8–11].

Despite increasing significance of communication skills training in a medical setting, there is a lack of a generally accepted definition of adequate physician-patient communication [12]. Based on five widely recognized physician-patient communication models, the Kalamazoo I Consensus Statement extracted a list of the following seven key elements that characterize adequate physician-patient communication: a) building relationship, b) opening discussion, c) gathering information, d) understanding the patient's perspective, e) sharing information, f) reaching agreement, and g) providing closure [9]. In addition, they represent a blueprint for the development of medical curricula comprising communication skills training and the assessment of students' performance [13,14]. Empirical studies have demonstrated the importance of a valid assessment of medical students' communication skills performance for several reasons [15]. First, through performance assessment students become aware of the relevance of physician-patient communication and receive feedback on their performance and deficits. Second, it enables educators to identify those medical students with significant deficits and reveals existing weaknesses within the curricula. Furthermore, summative assessments such as high-stake examinations could result in the denial of graduation in case of not qualified students to prevent damage from future patients [16].

To assess communication skills, most medical schools established the Objective Structured Clinical Examination (OSCE) using interactions with standardized patients (SP) [17]. An OSCE consists of several stations with different tasks and aims to simulate real clinical encounters between physician and patient. At that point it is important to emphasize that different kinds of OSCEs exist. They differ in their purpose. While some OSCEs address the assessment of communication skills in an integrated way as part of other clinical tasks (e.g. history taking, physical examination) there are also OSCEs which exclusively focus on the assessment of communication skills [2]. For the purpose of rating a student's communication skills performance

during an OSCE different kinds of rating scales have been developed [18–20]. Yet, the validity of the performance scores of a student is fundamentally dependent of the quality of the rating scales in use [21]. Nevertheless, a clear overview of the existing rating scales and their methodological and psychometric quality has not been conducted so far. Hence, a systematic review is needed to a) to compare and evaluate the existing rating scales based on well-defined quality criteria, b) to facilitate the choice of an appropriate instrument depending on the respective purpose, and c) to illustrate the gaps and needs in research, such as initiating the development of new instruments.

Therefore, this systematic review of rating scales on communication skills assessment in OSCEs aims at 1) identifying existing psychometrically tested rating scales on communication skills assessment in OSCEs and describing their underlying definition of communication skills, 2) determining the quality of design, methods and reporting of studies that analyze psychometric properties of rating scales, and 3) evaluating the psychometric quality of the identified rating scales.

## Methods

### Search strategy

We started our systematic review by performing an electronic literature search in the data bases EMBASE, PsycINFO and PubMed. We included all articles published between January 1979, the year in which the first OSCE to assess medical students' clinical competence was developed [22], and January 2, 2015. For this purpose, it was necessary to devise a specific search strategy for each of the three data bases based on a combination of different terms and keywords from the following four domains: (i) construct, (ii) context, (iii) measurement, and (iv) psychometric properties. In addition, we made use of the PubMed search filter developed by Terwee et al. [23] to facilitate the search process for studies on psychometric properties of rating scales. Based on our predefined inclusion and exclusion criteria, we limited each of the three specific search strategies to peer-reviewed publications, published in English or German. Furthermore, we also excluded studies in which communication skills were just reported as a subscale and thus did not allow the extraction of results related solely to this subscale. The applied inclusion and exclusion criteria are displayed in [Table 1](#). The full electronic search strategy is displayed in [S1 Appendix](#). As part of our search strategy, we also performed a secondary search which consisted of reference tracking of all included full texts and consultation of experts in the field of communication skills in health care.

### Study selection

First, we imported all search results into reference management software (EndNote) and removed all existing duplicates. Second, two reviewers (JZ and MC) independently performed a title and abstract screening to double-check the identified records for possible inclusion. In a next step, the remaining full texts were independently assessed for eligibility by two reviewers (EC and MC) using the inclusion and exclusion criteria. In case of disagreement regarding inclusion decisions, a third reviewer (IS) was consulted to reach consensus and to make a final decision.

### Data extraction and quality assessments

Final data extraction sheets were developed after pilot testing and adjustment in discussion between two reviewers (IS and MC). Data extraction sheets contained both descriptive data and data to assess the quality of the included studies. The process of assessing the quality

**Table 1. Inclusion and exclusion criteria.**

| <i>Inclusion criteria</i>   | <i>Excluded full texts<br/>(n = 49)</i> |
|---|---|
| 1 The article is published in a peer-reviewed journal                                 | 4                                       |
| 2 The language of the publication is English or German                                | 1                                       |
| 3 The publication date is between 1979 and 2015                                       |   |
| 4 The measured construct is communication skills                                      | 16                                      |
| 5 The examinee is a single medical student  | 4                                       |
| 6 The underlying rating scale is used in an OSCE with standardized patients           | 5                                       |
| 7 The underlying rating scale is exclusively used by examiner                         | 3                                       |
| 8 The rating scale is exclusively focused on communication skills                     | 4                                       |
| 9 The aim of the study is to test the psychometric properties of the underlying scale | 12                                      |
| <i>Exclusion criteria</i>   |   |
| 1 Not retrievable due to incomplete reference   |   |
| 2 Full text not available   |   |

Empty space = no full text was excluded for this reason.

doi:10.1371/journal.pone.0152717.t001

comprised two separate steps. As a first step, the quality of design, methods and reporting of the included studies on psychometric properties was assessed by applying the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist with 4-point scale [24–26]. The second step addressed the evaluation of the psychometric properties of the identified rating scales with the quality criteria developed by Terwee et al. [27]. The COSMIN checklist and the quality criteria for good psychometric properties developed by Terwee et al. are described below. To ensure consistency in the application of the COSMIN checklist and the quality criteria by Terwee et al., an independent double assessment (EC and MC) was performed for a random sample of 15% of included papers (i.e. two studies) at the start of data collection. Any eventual initial disagreements and ambiguities were resolved through discussion prior to extracting and rating data for the remaining 85% of studies. Finally, data extraction and quality assessment were conducted by one reviewer (MC).

**Assessment of methodological quality.** The COSMIN checklist was developed in a multi-disciplinary, international Delphi study and serves as a standardized tool for assessing the methodological quality of studies on measurement properties [24,25]. The COSMIN checklist consists of twelve boxes of which nine contain assessment standards for the following measurement properties: internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing, cross-cultural validity, criterion validity and responsiveness. In addition, according to the predetermined instructions for completing the COSMIN checklist, it is necessary to complete the IRT box if Item-Response-Theory methods were used in a study [26]. Furthermore, there are two boxes on interpretability and generalizability, which serve the purpose to extract descriptive data. The number of items of the boxes varies between five and eighteen. Each of these items can be scored on the 4-point scale as excellent (+++), good (++), fair (+), or poor (0) based on specific criteria. To obtain an overall score for a box, the lowest score of any item has to be taken, which is called the “worst score counts” method. While we performed data extraction and evaluation for each of the twelve COSMIN boxes, we omitted the presentation of the boxes interpretability and generalizability because they do not provide further information to our descriptive data extraction of the included studies. It should be mentioned that the COSMIN checklist was primarily developed to facilitate the assessment of the methodological quality of Health-Related Patient-Reported Outcomes (HR-PROs) [24].

Since this systematic review exclusively focuses on observer based rating scales to assess communication skills of medical students within an OSCE, some of the items of the COSMIN checklist were rated as “not applicable” (n/a).

**Assessment of psychometric quality.** The criteria developed by Terwee et al. [27] were used to assess the quality of the psychometric properties. They have been successfully applied in previous reviews [21,28,29], one of them also including observer measures [29]. The Terwee et al. criteria address the following properties: content validity, internal consistency, criterion validity, construct validity, reproducibility (agreement and reliability), responsiveness, floor and ceiling effects and interpretability. Each of those eight properties can be evaluated by one item as positive (+), intermediate (?), negative (-) or no information available (0).

## Results

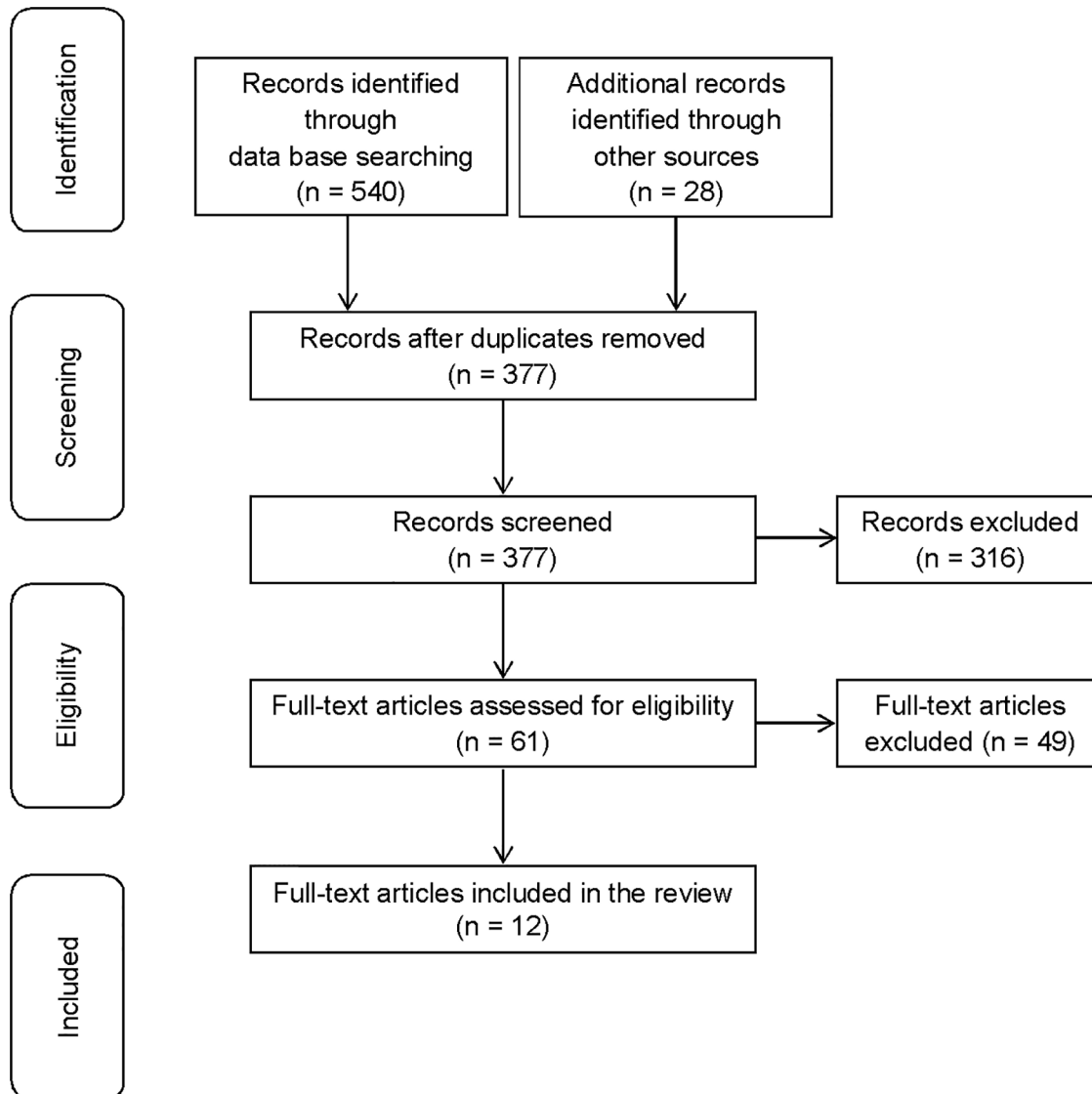
### Literature search and study selection

The electronic data base search yielded 540 records. In addition, 28 records were identified through secondary search of which 25 were from reference tracking and three from consultation of experts in the field of communication in health care. In a next step, 191 duplicates were removed. We then excluded another 316 records based on title and abstract screening. The full texts of the remaining 61 records were assessed for eligibility. Of the 61 records, 49 were excluded by applying the inclusion and exclusion criteria (see [Table 1](#)). As a result, twelve studies were included in this review. Most of the full texts were excluded either because the measured construct was not communication skills (n = 16) or the aim of the study was not to test the psychometric properties (n = 12). The study selection procedure is shown in [Fig 1](#).

### Description of included studies and rating scales

The majority of the included studies were conducted in Europe. Of the twelve included studies reporting on eight rating scales, five were from UK [30–34], three from Germany [35–37], two from Canada [38,39] and one each from Belgium [40] and the US [41]. The study samples exclusively consisted of undergraduate medical students, with two of the studies being carried out during clinical clerkship [38,39]. Seven studies were initial studies with the objective of examining psychometric properties of a new measure [31–35,39,41]. The other five studies were further examinations of previously developed rating scales and reporting on their psychometric properties [30, 36–38,40]. Looking at the setting of the studies, it is important to underline that the OSCEs differed in their purpose between formative and summative evaluations. While formative OSCEs provide the examinee with performance feedback, summative OSCEs enable the examiners to make pass-fail decisions based on predefined criteria [42,43]. Most of the OSCEs in our systematic review were exclusively used for summative evaluations [30,31, 33, 34,36,38,40]. Descriptive data of the included studies are displayed in [Table 2](#).

The present review included eight rating scales which have been applied in OSCE settings to assess communication skills of medical students while they interacted with SPs. From these eight rating scales, five were clearly named by the authors [30–33, 38,40,41]. For the remaining three rating scales we had come up with an acronym based on information from title or abstract. Thus, MCS-OSCE stands for the Mayence Communication Skills OSCE [35], AG-OSCE-R for the Analytic Global OSCE Rating [39] and finally LIDM-RS for the Leeds Informed Decision Making Rating Scale [34]. One of the three aims of this review was to describe the underlying definition of communication skills of the included rating scales. As displayed in [Table 3](#), not all of the eight rating scales are explicitly developed based on a clear definition of communication skills. Of the eight rating scales, five include a definition of communication skills [30,32,33,35,38,40,41]. The underlying definition of two rating scales [30,33,38] is based



**Fig 1. Flow diagram of study selection.**

doi:10.1371/journal.pone.0152717.g001

on the Calgary-Cambridge Guide, which is a model for medical interview [44,45]. One measure [40,41] derives its definition of communication skills from the Toronto and Kalamazoo Consensus Statements [9, 10]. Finally, there are two rating scales that contain their own specific definition of communication skills [32,35]. Descriptive data of the included rating scales are shown in Table 3.

### Quality of design, methods and reporting

The assessment of the methodological quality of the included studies on measurement properties by applying the COSMIN checklist is presented in Table 4. None of the twelve studies reported on all of the nine COSMIN boxes. One of the twelve included studies used Item-Response-Theory [38]. Furthermore, another three studies applied the generalizability theory



**Table 2. Descriptive data of the included studies.**

| Measure   | Authors (Year)           | Setting*   | Study sample*  | Country |
|-----------|--------------------------|--|--|---------|
| EPSCALE   | Silverman et al. (2011)  | summative OSCE with UMS during final MB examinations at University of Cambridge  | n = 124**  | UK      |
| EPSCALE   | Edgcumbe et al. (2012)   | summative OSCE with UMS during final MB examinations at University of Cambridge  | n = 124**  | UK      |
| MCS-OSCE  | Fischbeck et al. (2011)* | formative and summative OSCE with UMS attending a Medical Psychology and Medical Sociology course at Johannes Gutenberg University Mainz | sample 1: n = 182, 61% f, m <sub>age</sub> = 22 ys; sample 2: n = 181, 66% f, m <sub>age</sub> = 22 ys | Germany |
| CCAT      | Harasym et al. (2008)    | summative OSCE with UMS during clinical clerkship at University of Calgary   | n = 190  | Canada  |
| AG-OSCE-R | Hodges & McIlroy (2003)  | formative OSCE with UMS during clinical clerkship in 3rd or 4th study year at University of Toronto                                      | n = 57   | Canada  |
| AG-OSCE-R | Scheffer et al. (2008)   | formative OSCE with UMS in 2nd or 3rd study year at Charité-Universitätsmedizin Berlin   | n = 113, 65% f, m <sub>age</sub> = 23 ys (SD <sub>age</sub> = 4)                                       | Germany |
| AG-OSCE-R | Mortsiefer et al. (2014) | summative OSCE with UMS in 4th study year at Heinrich-Heine-University Düsseldorf  | n = 453  | Germany |
| LCSAS     | Humphris & Kaney (2001)* | summative OSCE with UMS in 1st or 2nd study year at University of Liverpool  | sample 1: n = 600; sample 2: n = 60; sample 3: n = 80  | UK      |
| LUCAS     | Huntley et al. (2012)*   | formative and summative OSCE with UMS in 1st study year at University of Liverpool   | sample 1: n = 731, 53.7% f, m <sub>age</sub> = 19 ys (SD <sub>age</sub> = 1.61); sample 2: n = 40      | UK      |
| LIDM-RS   | Thistlethwaite (2002)    | summative OSCE with UMS in 3rd study year at University of Leeds   | n = 194  | UK      |
| CG        | Lang et al. (2004)*      | formative OSCE with UMS of two cohorts at East Tennessee State University and Tulane University  | sample 1: n = 50; sample 2: n = 50   | USA     |
| CG        | Van Nuland et al. (2012) | summative OSCE with UMS in their final undergraduate year at Catholic University of Leuven   | n = 63   | Belgium |

\*some studies were conducted in more than one setting and/or used more than one sample,

\*\*Silverman et al. (2011) and Edgcumbe et al. (2012) used the same study sample, UMS = undergraduate medical students, MB = Medicinae Baccalaureus,

f = female, ys = years, m = mean, SD = standard deviation. Full titles of the rating scales: Explanation and Planning Scale (EPSCALE), Mayence Communication Skills OSCE (MCS-OSCE), Calgary-Cambridge Assessment Tool (CCAT), Analytic Global OSCE Rating (AG-OSCE-R), Liverpool Communication Skills Assessment Scale (LCSAS), Liverpool Undergraduate Communication Assessment Scale (LUCAS), Leeds Informed Decision Making Rating Scale (LIDM-RS), Common Ground (CG).

doi:10.1371/journal.pone.0152717.t002

[46] to assess reliability (Box B) [31,33,40]. To assess these studies properly, it was necessary to make minor adjustments to the respective COSMIN box. Internal consistency (Box A) was calculated in seven studies [32,33,35,36,39–41]. Only one of them [32] received an excellent score, while the other six studies [33,35,36,39–41] were rated poor. Reliability (Box B) was addressed in ten studies [31–38,40,41]. Six of them scored excellent [31–33,36–38], one good [40], one fair [35] and two poor [34,41]. Measurement error (Box C) was not reported in any of the included studies. Of the seven studies [31–35,39,41], where content validity could be rated (Box D), only one study was rated fair [31], while the other six studies scored poorly [32–35,39,41]. Only two studies addressed the structural validity (Box E) [30,32] of which one scored excellent [32] and one good [30]. Hypotheses testing (Box F) was conducted in eight studies [31,32,35–37,39–41]. Two of them [32,35] were rated good, whereas the other six studies [31,36,37,39–41] received a poor score. One study [37] translated a measure into German and received a poor score regarding the translation procedure. Detailed results for COSMIN ratings on item level are shown in the [S2 Appendix](#).

**Table 3. Descriptive data of the included rating scales.**

| Measure          | Authors (Year)           | Underlying definition of communication skills                  | Language | Dimensions | Items | Response                            |
|------------------|--------------------------|--|----------|------------|-------|-------------------------------------|
| <b>EPSCALE</b>   | Silverman et al. (2011)  | definition based on Calgary-Cambridge Guide                    | English  | 6          | 15    | 4-point scale                       |
|                  | Edgcumbe et al. (2012)   | definition based on Calgary-Cambridge Guide                    | English  | 6          | 15    | 4-point scale                       |
| <b>MCS-OSCE</b>  | Fischbeck et al. (2011)  | own definition   | German   | n/r        | n/r   | 5-point scale                       |
| <b>CCAT</b>      | Harasym et al. (2008)    | definition based on Calgary-Cambridge Guide                    | English  | 3          | 28    | 3-point scale                       |
| <b>AG-OSCE-R</b> | Hodges & McIlroy (2003)  | n/r  | English  | 4          | 4     | 5-point scale                       |
|                  | Scheffer et al. (2008)   | n/r  | German   | 4          | 4     | 5-point scale                       |
|                  | Mortsiefer et al. (2014) | n/r  | German   | 4          | 4     | 5-point scale                       |
| <b>LCSAS*</b>    | Humphris & Kaney (2001)  | n/r  | English  | 5          | 12    | 4-point scale                       |
| <b>LUCAS</b>     | Huntley et al. (2012)    | own definition   | English  | 2          | 10    | 4 items using 2 response options**; |
|                  |                          |  |          |            |       | 6 items using 3 response options*** |
| <b>LIDM-RS</b>   | Thistlethwaite (2002)    | n/r  | English  | n/r        | 10    | 3-point scale                       |
| <b>CG</b>        | Lang et al. (2004)       | definition based on Toronto and Kalamazoo Consensus Statements | English  | 7          | 36    | mixed                               |
|                  | Van Nuland et al. (2012) | definition based on Toronto and Kalamazoo Consensus Statements | Dutch    | 7          | 7     | 7-point scale                       |

\*The LCSAS is just one part of the communication skills assessment system of Humphris & Kaney (2001). The LCSAS is intended to be applied with the Global Simulated Patient Rating Scale (GSPRS). We excluded the GSPRS since it is a rating scale used by SPs.

\*\*The two response options were competent-unacceptable.

\*\*\*The three response options were competent-borderline-unacceptable. n/r = not reported, cs = communication skills. Full titles of the rating scales: Explanation and Planning Scale (EPSCALE), Mayence Communication Skills OSCE (MCS-OSCE), Calgary-Cambridge Assessment Tool (CCAT), Analytic Global OSCE Rating (AG-OSCE-R), Liverpool Communication Skills Assessment Scale (LCSAS), Liverpool Undergraduate Communication Assessment Scale (LUCAS), Leeds Informed Decision Making Rating Scale (LIDM-RS), Common Ground (CG).

doi:10.1371/journal.pone.0152717.t003

### Quality of psychometric properties

The evaluation of the psychometric properties of the identified rating scales was carried out by applying the quality criteria of Terwee et al. The corresponding results are shown in [Table 5](#). Content validity received negative scores in all of the seven respective studies [[31–35,39,41](#)]. In case of internal consistency a positive rating was received in one study [[32](#)] and an intermediate rating in six studies [[31,32,39–41](#)]. Construct validity was evaluated positively in three studies [[35–37](#)] and received intermediate rating in five studies [[31,32,39–41](#)]. None of the included studies provided any information on agreement, while ten studies contained information on reliability. Reliability was rated positively in three studies [[32,37,38](#)], intermediate in one study [[41](#)], and negative in six studies [[31,33–36,40](#)]. Regarding criterion validity, responsiveness and floor and ceiling effects, none of the studies gave any information. Finally, interpretability was reported in seven studies and was judged as intermediate in all of them [[31,34–37,39,40](#)].



**Table 4. Quality of design, methods and reporting of studies on psychometric properties.**

| Measure   | Authors (Year)           | IRT Box | Psychometric properties |                  |   |   |   |     |    |                |   |  |
|-----------|--------------------------|---------|-------------------------|------------------|---|---|---|-----|----|----------------|---|--|
|           |                          |         | A                       | B                | C | D | E | F   | G  | H              | I |  |
| EPSCALE   | Silverman et al. (2011)  |         | 0                       | +++              |   | 0 |   |     |    |                |   |  |
|           | Edgcumbe et al. (2012)   |         |                         |                  |   |   |   | ++  |    |                |   |  |
| MCS-OSCE  | Fischbeck et al. (2011)  |         | 0                       | + <sup>a</sup>   |   | 0 |   |     | ++ |                |   |  |
| CCAT      | Harasym et al. (2008)    | +++     |                         | +++              |   |   |   |     |    |                |   |  |
| AG-OSCE-R | Hodges & McIlroy (2003)  |         | 0                       |                  |   | 0 |   |     | 0  |                |   |  |
|           | Scheffer et al. (2008)   |         |                         | +++ <sup>a</sup> |   |   |   |     | 0  | 0 <sup>c</sup> |   |  |
|           | Mortsiefer et al. (2014) |         | 0                       | +++ <sup>a</sup> |   |   |   |     | 0  |                |   |  |
| LCSAS     | Humphris & Kaney (2001)  |         |                         | +++ <sup>a</sup> |   | + |   |     | 0  |                |   |  |
| LUCAS     | Huntley et al. (2012)    |         | +++                     | +++ <sup>a</sup> |   | 0 |   | +++ | ++ |                |   |  |
| LIDM-RS   | Thistlethwaite (2002)    |         |                         | 0 <sup>a</sup>   |   | 0 |   |     |    |                |   |  |
| CG        | Lang et al. (2004)       |         | 0                       | 0 <sup>a,b</sup> |   | 0 |   |     | 0  |                |   |  |
|           | Van Nuland et al. (2012) |         | 0                       | ++               |   |   |   |     | 0  |                |   |  |

COSMIN psychometric property boxes: IRT Box = General requirements for studies that applied Item Response Theory (IRT) models, A = internal consistency, B = reliability, C = measurement error, D = content validity, E = structural validity, F = hypotheses testing, G = cross-cultural validity, H = criterion validity, I = responsiveness. 4-point scale rating: +++ = excellent, ++ = good, + = fair, 0 = poor, empty space = COSMIN rating not applicable. For exact information regarding the definitions of psychometric properties and 4-point scale rating see COSMIN website ([www.cosmin.nl](http://www.cosmin.nl)).

<sup>a</sup> = Inter-rater-reliability,

<sup>b</sup> = Intra-rater-reliability,

<sup>c</sup> = only evaluation of the quality of the translation procedure.

doi:10.1371/journal.pone.0152717.t004

## Discussion

The present systematic review aimed at identifying psychometrically tested rating scales on communication skills assessment in OSCEs, describing their underlying definition of communication skills, assessing the methodological quality of the included studies and evaluating the psychometric quality of the identified rating scales. For these purposes, data were extracted from twelve studies reporting on eight rating scales.

Regarding the underlying definition of communication skills of the identified rating scales, publications on three of the eight identified rating scales (AG-OSCE-R [36,37,39], LCSAS [31] and LIDM-RS [34]) did not provide any information on how communication skills were defined. This is certainly a shortcoming, as it would be important for readers of these papers to know on what basis items were developed, especially for educators, who might want to use these scales for OSCE assessment at their university. On the other hand, many of the rating scales (EPSCALE [30,33], CCAT [38] and CG [40,41]) are either based on the well-known model of Calgary-Cambridge Guide [44,45] or on the much-cited consensus statements of Toronto and Kalamazoo [9, 10]. In terms of using one of the identified rating scales in a specific medical education setting, we recommend checking whether a measure's definition of communication skills matches the definition given in the curriculum of the specific setting.

The process of assessing the methodological quality of the included studies by applying the COSMIN checklist revealed that most studies were mainly poorly rated. One exception was the quality of the assessment of reliability, which was rated as excellent in most studies. Another main exception was the study reporting on psychometric properties of LUCAS [32], which received mainly excellent and good scores. However, its content validity was rated of poor quality. Another study worth mentioning positively was the one reporting on psychometric properties of the CCAT [38]. Although it only tested reliability by using the Item-Response-Theory, it

**Table 5. Quality of psychometric properties.**

| <i>Instruments / Authors / (Year)</i> | <i>Content validity</i> | <i>Internal consistency</i> | <i>Criterion validity</i> | <i>Construct validity</i> | <i>Reproducibility (Agreement)</i> | <i>Reproducibility (Reliability)</i> | <i>Responsiveness</i> | <i>Floor &amp; ceiling effects</i> | <i>Interpretability</i> |
|---------------------------------------|-------------------------|-----------------------------|---------------------------|---------------------------|------------------------------------|--------------------------------------|-----------------------|------------------------------------|-------------------------|
| <b>EPSCALE</b>                        | -0                      | ?0                          | 00                        | 00                        | 00                                 | -0                                   | 00                    | 00                                 | 00                      |
| Silverman et al. (2011)               | -                       | ?                           | 0                         | 0                         | 0                                  | -                                    | 0                     | 0                                  | 0                       |
| Edgcumbe et al. (2012)                | 0                       | 0                           | 0                         | 0                         | 0                                  | 0                                    | 0                     | 0                                  | 0                       |
| <b>MCS-OSCE</b>                       | -                       | ?                           | 0                         | +                         | 0                                  | -                                    | 0                     | 0                                  | ?                       |
| Fischbeck et al. (2011)               | -                       | ?                           | 0                         | +                         | 0                                  | -                                    | 0                     | 0                                  | ?                       |
| <b>CCAT</b>                           | 0                       | 0                           | 0                         | 0                         | 0                                  | +                                    | 0                     | 0                                  | 0                       |
| Harasym et al. (2008)                 | 0                       | 0                           | 0                         | 0                         | 0                                  | +                                    | 0                     | 0                                  | 0                       |
| <b>AG-OSCE-R</b>                      | -00                     | ?0?                         | 000                       | ?++                       | 000                                | 0+-                                  | 000                   | 000                                | ???                     |
| Hodges & Mclroy (2003)                | -                       | ?                           | 0                         | ?                         | 0                                  | 0                                    | 0                     | 0                                  | ?                       |
| Scheffer et al. (2008)                | 0                       | 0                           | 0                         | +                         | 0                                  | +                                    | 0                     | 0                                  | ?                       |
| Mortsiefer et al. (2014)              | 0                       | ?                           | 0                         | +                         | 0                                  | -                                    | 0                     | 0                                  | ?                       |
| <b>LCSAS</b>                          | -                       | 0                           | 0                         | ?                         | 0                                  | -                                    | 0                     | 0                                  | ?                       |
| Humphris & Kaney (2001)               | -                       | 0                           | 0                         | ?                         | 0                                  | -                                    | 0                     | 0                                  | ?                       |
| <b>LUCAS</b>                          | -                       | +                           | 0                         | ?                         | 0                                  | +                                    | 0                     | 0                                  | 0                       |
| Huntley et al. (2012)                 | -                       | +                           | 0                         | ?                         | 0                                  | +                                    | 0                     | 0                                  | 0                       |
| <b>LIDM-RS</b>                        | -                       | 0                           | 0                         | 0                         | 0                                  | -                                    | 0                     | 0                                  | ?                       |
| Thistlethwaite (2002)                 | -                       | 0                           | 0                         | 0                         | 0                                  | -                                    | 0                     | 0                                  | ?                       |
| <b>CG</b>                             | -0                      | ??                          | 00                        | ??                        | 00                                 | ?-                                   | 00                    | 00                                 | 0?                      |
| Lang et al. (2004)                    | -                       | ?                           | 0                         | ?                         | 0                                  | ?                                    | 0                     | 0                                  | 0                       |
| Van Nuland et al. (2012)              | 0                       | ?                           | 0                         | ?                         | 0                                  | -                                    | 0                     | 0                                  | ?                       |

Rating: + = positive, ? = intermediate,— = negative, 0 = no information available. Grey lines summarize ratings of psychometric properties per measure. For exact information regarding the definitions of psychometric properties see Terwee et al. [27].

doi:10.1371/journal.pone.0152717.t005

was rated of excellent quality. When comparing the COSMIN ratings between studies, the measure development study of the CG [41] received the lowest ratings. All of four psychometric properties reported in this study were rated poor. Looking at the COSMIN ratings on the item level (see S2 Appendix), it is important to emphasize that they reveal a more differentiated picture. Several studies scored excellent or good on many items of the nine COSMIN boxes. However, under the terms of the “worst score counts” method of COSMIN to obtain an overall score for a box the lowest score of any item had to be taken, which led to poor psychometric property ratings for many studies. Thus, many studies could have performed much better in terms of methodological quality, if they would have taken into account the recommendations of the COSMIN group.

The evaluation of the psychometric properties using the criteria developed by Terwee et al. showed that the psychometric quality of the eight identified rating scales was mainly

intermediate. The measure LUCAS [32] received the best rating in terms of psychometric quality. However, it is remarkable that none of the rating scales received a positive or an intermediate quality rating on content validity. Based on the fact that content validity is meant to be one of the most important psychometric properties [27], these assessments on content validity represent a major flaw.

The corresponding results of the methodological quality resulting from the COSMIN checklist and of the psychometric quality with the Terwee et al. criteria have to be taken into account together to draw conclusions appropriately. In this review several serious flaws concerning design, methods and reporting of the included studies could be shown by applying the COSMIN checklist. Thus, it is important to note that the results of the Terwee criteria on the psychometric quality of the rating scales need to be interpreted with care, as it is difficult to say how much one can trust the results gained from studies with poor design, methods and reporting. Combining the results of the COSMIN checklist and the Terwee et al. criteria, LUCAS [32] had the best results. Nevertheless, it must be underlined that its content validity is not satisfactory and should be checked in future research. It is also important to mention that some of the rating scales scored excellent or good on the methodical rating with COSMIN, while the evaluation with the Terwee et al. criteria clearly revealed poor psychometric properties. These results have a higher credibility than those gained from methodologically flawed studies.

Our systematic review has several strengths. First, we devised a specific search strategy for each of the three data bases in order to identify all records relevant to our purpose. Second, two reviewers independently performed a title and abstract screening to double-check the identified records for possible inclusion. Third, as recommended, the process of assessing the quality comprised two separated steps using the COSMIN checklist with 4-point scale rating to rate the methodological quality of the included studies and the quality criteria for good psychometric properties developed by Terwee et al. to determine the quality of the psychometric properties. The assessment of the methodological quality of the included studies is intended to make sure that psychometric properties reported in the studies can be interpreted and rated appropriately. Besides its strengths, this present review has also several limitations. First, our search was limited to English and German. Hence, it is possible that we might have failed to notice relevant publications. To minimize this risk, we also performed a secondary search which consisted of reference tracking of all included full texts and consultation of a range of international experts in the field of communication in health care. Second, 85% of the process of data extraction and quality assessment was performed by one reviewer only. Thus, it cannot be excluded with certainty that the assessment of included studies and psychometric quality of the identified rating scales were biased. However, a double assessment was performed for the first two studies in order to discuss and to resolve eventual initial ambiguities regarding the application of the COSMIN checklist and the Terwee et al. criteria. Third, due to our inclusion and exclusion criteria, we exclusively focused on rating scales used by examiners. Thus, we excluded rating scales that are meant to be completed by standardized patients to assess medical students' communication skills. These tools can also be of high value, especially for formative assessment of communication skills and it might be interesting for future research to examine the performance of those measures as well.

In this systematic review eight rating scales assessing the communication skills of medical students in OSCE settings were identified. According to our results, the development of new rating scales is not necessarily required. Instead, efforts need to be made to eliminate the existing flaws. The COSMIN checklist illustrated several research gaps in the methodological quality of psychometric evaluation studies, which have to be approached. Since the methodological quality of the psychometric evaluation studies represents the basis for the evaluation of psychometric properties, it is indispensable to improve it. For this purpose, we recommend to use

more rigorous methodological designs and a more detailed reporting. First, future psychometric studies need to conduct and describe the testing of content validity in more detail. Second, analyses of the factorial structure of the rating scales should be performed, which has an impact on internal consistency and structural validity. For hypotheses testing (on convergent or divergent validity) to be improved, future evaluation studies need clearly formulated hypotheses, larger sample sizes for multiple hypotheses and an adequate description of the comparator rating scales. Third, several psychometric properties (e.g. measurement error, floor and ceiling effects, responsiveness) were completely neglected in all included studies. Thus, they deserve attention in future psychometric evaluation studies.

## Conclusion

Our systematic review gives an overview of rating scales, which are applied within the medical education setting to assess students' communication skills. It can help teachers and researchers in the field of medical education to find the appropriate measure for their individual purpose. Nevertheless, we identified several research gaps regarding the methodological quality of studies reporting on psychometric properties and the quality of their results. Based on our results, the use of the eight identified rating scales to assess students' communication skills needs to be done with care, as their methodological quality is not completely satisfactory. Hence, future psychometric evaluation studies focusing on improving the methodological quality are needed in order to yield psychometrically sound results of the OSCEs assessing communication skills. This is especially important considering that most rating scales included in this review were used for summative evaluation, i.e. to make pass-fail decisions. Such decisions have a high impact on students' academic success and should be based on reliable and valid assessment.

## Supporting Information

**S1 Appendix. Electronic data base search strategy for EMBASE, PsycINFO, PubMed.**  
(DOCX)

**S2 Appendix. Detailed results for the COSMIN checklist.**  
(DOCX)

**S1 Checklist. PRISMA checklist.**  
(DOC)

## Acknowledgments

The authors wish to thank Levente Kriston for his advice on quality assessment and the members of the Research Committee of the European Association of Communication in Health Care (rEACH) for their expert feedback as part of the secondary search strategy.

## Author Contributions

Conceived and designed the experiments: MC JZ JD MH IS. Performed the experiments: MC JZ EC IS. Analyzed the data: MC JZ EC IS. Wrote the paper: MC JZ EC JD MH IS.

## References

1. Hausberg MC, Hergert A, Kröger C, Bullinger M, Rose M, Andreas S (2012) Enhancing medical students' communication skills: Development and evaluation of an undergraduate training program. *BMC Medical Education* 12:
2. Baig LA, Violato C, Crutcher RA (2009) Assessing clinical communication skills in physicians: Are the skills context specific or generalizable. *BMC Medical Education* 9:

3. Accreditation Council for Graduate Medical Education (ACGME) (2007) Common program requirements. Available: [https://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/CPRs\\_07012015.pdf](https://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/CPRs_07012015.pdf) Accessed 15 Oct 2015.
4. American Board of Medical Specialties (ABMS) (2014) Standards for the ABMS Program for Maintenance of Certification (MOC). Available: <http://www.abms.org/media/1109/standards-for-the-abms-program-for-moc-final.pdf> Accessed 15 Oct 2015.
5. Association of American Medical Colleges (AAMC) (1998) Report I. Learning objectives for medical student education. Guidelines for medical schools. Available: <https://members.aamc.org/eweb/upload/Learning%20Objectives%20for%20Medical%20Student%20Educ%20Report%20I.pdf> Accessed 16 Oct 2015.
6. General Medical Council (GMC) (2009) Tomorrow's doctors. Outcomes and standards for undergraduate medical education. Available: [http://www.gmc-uk.org/Tomorrow\\_s\\_Doctors\\_1214.pdf](http://www.gmc-uk.org/Tomorrow_s_Doctors_1214.pdf) 48905759.pdf Accessed 15 Oct 2015.
7. World Federation for Medical Education (WFME) (2015) Basic medical education. WFME global standards for quality improvement. The 2015 revision. Available: <http://wfme.org/standards/bme/78-new-version-2012-quality-improvement-in-basic-medical-education-english/file> Accessed 15 Oct 2015.
8. Kiessling C, Dieterich A, Fabry G, Hölzer H, Langewitz W, Mühlinghaus I, et al. (2008) Basel Consensus Statement "Communicative and Social Competencies in Medical Education": A position paper of the GMA Committee Communicative and Social Competencies [German]. *GMS Zeitschrift für Medizinische Ausbildung* 25: Doc83 (20080515)
9. Makoul G (2001) Essential elements of communication in medical encounters: The Kalamazoo Consensus Statement. *Academic Medicine* 76: 390–393. PMID: [11299158](#)
10. Simpson M, Buckman R, Stewart M, Maguire P, Lipkin M, Novack D, et al. (1991) Doctor-patient communication: The Toronto Consensus Statement. *BMJ: British Medical Journal* 303: 1385–1387. PMID: [1760608](#)
11. Von Fragstein M, Silverman J, Cushing A, Quilligan S, Salisbury H, Wiskin C, et al. (2008) UK Consensus Statement on the content of communication curricula in undergraduate medical education. *Medical Education* 42: 1100–1107. doi: [10.1111/j.1365-2923.2008.03137.x](https://doi.org/10.1111/j.1365-2923.2008.03137.x) PMID: [18761615](#)
12. Deveugele M, Derese A, Maesschalck SD, Willems S, Van Driel M, De Maeseneer J (2005) Teaching communication skills to medical students, a challenge in the curriculum? *Patient Education and Counseling* 58: 265–270. PMID: [16023822](#)
13. Duffy FD, Gordon GH, Whelan G, Cole-Kelly K, Frankel R, All Participants in the American Academy on Physician and Patient's Conference on Education and Evaluation of Competence in Communication and Interpersonal Skills (2004) Assessing competence in communication and interpersonal skills: The Kalamazoo II Report. *Academic Medicine* 79: 495–507. PMID: [15165967](#)
14. Joyce BL, Steenbergh T, Scher E (2010) Use of the Kalamazoo Essential Elements Communication Checklist (Adapted) in an institutional interpersonal and communication skills curriculum. *Journal of Graduate Medical Education* 2: 165–169. doi: [10.4300/JGME-D-10-00024.1](https://doi.org/10.4300/JGME-D-10-00024.1) PMID: [21975614](#)
15. Bergus GR, Kreiter CD (2007) The reliability of summative judgements based on objective structured clinical examination cases distributed across the clinical year. *Medical Education* 41: 661–666. PMID: [17614886](#)
16. De Haes JCJM, Oort FJ, Hulsman RL (2005) Summative assessment of medical students' communication skills and professional attitudes through observation in clinical practice. *Medical Teacher* 27: 583–589. PMID: [16332548](#)
17. Bergus GR, Woodhead JC, Kreiter CD (2009) Trained lay observers can reliably assess medical students' communication skills. *Medical Education* 43: 688–694. doi: [10.1111/j.1365-2923.2009.03396.x](https://doi.org/10.1111/j.1365-2923.2009.03396.x) PMID: [19573193](#)
18. Newble D (2004) Techniques for measuring clinical competence: Objective structured clinical examinations. *Medical Education* 38: 199–203. PMID: [14871390](#)
19. Regehr G, MacRae H, Reznick RK, Szalay D (1998) Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine* 73: 993–997. PMID: [9759104](#)
20. Schirmer JM, Mauksch L, Lang F, Marvel MK, Zoppi K, Epstein RM, et al. (2005) Assessing communication competence: A review of current tools. *Family Medicine* 37: 184–192. PMID: [15739134](#)
21. Müller E, Zill JM, Dirmaier J, Härter M, Scholl I (2014) Assessment of trust in physician: A systematic review of measures. *PLoS ONE* 9: e106844. doi: [10.1371/journal.pone.0106844](https://doi.org/10.1371/journal.pone.0106844) PMID: [25208074](#)
22. Harden RM, Gleeson FA (1979) Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education* 13: 41–54. PMID: [763183](#)

23. Terwee CB, Jansma EP, Riphagen II, de Vet HCW (2009) Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research* 18: 1115–1123. doi: [10.1007/s11136-009-9528-5](https://doi.org/10.1007/s11136-009-9528-5) PMID: [19711195](https://pubmed.ncbi.nlm.nih.gov/19711195/)
24. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research* 19: 539–549. doi: [10.1007/s11136-010-9606-8](https://doi.org/10.1007/s11136-010-9606-8) PMID: [20169472](https://pubmed.ncbi.nlm.nih.gov/20169472/)
25. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. (2010) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology* 63: 737–745. doi: [10.1016/j.jclinepi.2010.02.006](https://doi.org/10.1016/j.jclinepi.2010.02.006) PMID: [20494804](https://pubmed.ncbi.nlm.nih.gov/20494804/)
26. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research* 21: 651–657. doi: [10.1007/s11136-011-9960-1](https://doi.org/10.1007/s11136-011-9960-1) PMID: [21732199](https://pubmed.ncbi.nlm.nih.gov/21732199/)
27. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* 60: 34–42. PMID: [17161752](https://pubmed.ncbi.nlm.nih.gov/17161752/)
28. Tijssen M, van Cingel R, van Melick N, de Visser E (2011) Patient-Reported Outcome questionnaires for hip arthroscopy: A systematic review of the psychometric evidence. *BMC Musculoskeletal Disorders* 12:
29. Zill JM, Christalle E, Müller E, Härter M, Dirmaier J, Scholl I (2014) Measurement of physician-patient communication—A systematic review. *PLOS ONE* 9: doi: [10.1371/journal.pone.0112637](https://doi.org/10.1371/journal.pone.0112637)
30. Edgcumbe DP, Silverman J, Benson J (2012) An examination of the validity of EPSCALE using factor analysis. *Patient Education and Counseling* 87: 120–124. doi: [10.1016/j.pec.2011.07.011](https://doi.org/10.1016/j.pec.2011.07.011) PMID: [21852064](https://pubmed.ncbi.nlm.nih.gov/21852064/)
31. Humphris GM, Kaney S (2001) The Liverpool brief assessment system for communication skills in the making of doctors. *Advances in Health Sciences Education: Theory and Practice* 6: 69–80.
32. Huntley CD, Salmon P, Fisher PL, Fletcher I, Young B (2012) LUCAS: A theoretically informed instrument to assess clinical communication in objective structured clinical examinations. *Medical Education* 46: 267–276. doi: [10.1111/j.1365-2923.2011.04162.x](https://doi.org/10.1111/j.1365-2923.2011.04162.x) PMID: [22324526](https://pubmed.ncbi.nlm.nih.gov/22324526/)
33. Silverman J, Archer J, Gillard S, Howells R, Benson J (2011) Initial evaluation of EPSCALE, a rating scale that assesses the process of explanation and planning in the medical interview. *Patient Education and Counseling* 82: 89–93. doi: [10.1016/j.pec.2010.02.022](https://doi.org/10.1016/j.pec.2010.02.022) PMID: [20338713](https://pubmed.ncbi.nlm.nih.gov/20338713/)
34. Thistlethwaite JE (2002) 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. *Education for Health* 15: 170–179. PMID: [14741966](https://pubmed.ncbi.nlm.nih.gov/14741966/)
35. Fischbeck S, Mauch M, Leschnik E, Beutel ME, Laubach W (2011) Assessment of communication skills with an OSCE among first year medical students [German]. *Psychotherapie, Psychosomatik, Medizinische Psychologie* 61: 465–471. doi: [10.1055/s-0031-1291277](https://doi.org/10.1055/s-0031-1291277) PMID: [22081465](https://pubmed.ncbi.nlm.nih.gov/22081465/)
36. Mortsiefer A, Immecke J, Rotthoff T, Karger A, Schmelzer R, Raski B, et al. (2014) Summative assessment of undergraduates' communication competence in challenging doctor-patient encounters. Evaluation of the Düsseldorf CoMed-OSCE. *Patient Education and Counseling* 95: 348–355. doi: [10.1016/j.pec.2014.02.009](https://doi.org/10.1016/j.pec.2014.02.009) PMID: [24637164](https://pubmed.ncbi.nlm.nih.gov/24637164/)
37. Scheffer S, Muehlinghaus I, Froehmel A, Ortwein H (2008) Assessing students' communication skills: Validation of a global rating. *Advances in Health Sciences Education: Theory and Practice* 13: 583–592.
38. Harasym PH, Woloschuk W, Cuning L (2008) Undesired variance due to examiner stringency/leniency effect in communication skill scores assessed in OSCEs. *Advances in Health Sciences Education* 13: 617–632. PMID: [17610034](https://pubmed.ncbi.nlm.nih.gov/17610034/)
39. Hodges B, McIlroy JH (2003) Analytic global OSCE ratings are sensitive to level of training. *Medical Education* 37: 1012–1016. PMID: [14629415](https://pubmed.ncbi.nlm.nih.gov/14629415/)
40. Van Nuland M, Van den Noortgate W, van der Vleuten C, Goedhuys J (2012) Optimizing the utility of communication OSCEs: Omit station-specific checklists and provide students with narrative feedback. *Patient Education and Counseling* 88: 106–112. doi: [10.1016/j.pec.2011.12.015](https://doi.org/10.1016/j.pec.2011.12.015) PMID: [22322068](https://pubmed.ncbi.nlm.nih.gov/22322068/)
41. Lang F, McCord R, Harvill L, Anderson DS (2004) Communication assessment using the Common Ground Instrument: Psychometric properties. *Family Medicine* 36: 189–198. PMID: [14999576](https://pubmed.ncbi.nlm.nih.gov/14999576/)
42. Epstein RM (2007) Assessment in medical education. *New England Journal of Medicine* 356: 387–396. PMID: [17251535](https://pubmed.ncbi.nlm.nih.gov/17251535/)



43. Gormley G (2011) Summative OSCEs in undergraduate medical education. *The Ulster Medical Journal* 80: 127–132. PMID: [23526843](#)
44. Kurtz S, Silverman J, Benson J, Draper J (2003) Marrying content and process in clinical method teaching: Enhancing the Calgary—Cambridge Guides. *Academic Medicine* 78: 802–809. PMID: [12915371](#)
45. Kurtz SM, Silverman J, Draper J (1998) *Teaching and learning communication skills*. Oxford: Radcliffe Medical Press.
46. Crossley J, Davies H, Humphris G, Jolly B (2002) Generalisability: A key to unlock professional assessment. *Medical Education* 36: 972–978. PMID: [12390466](#)