The Utility of Mini-Clinical Evaluation Exercise (Mini-CEX) in Undergraduate and Postgraduate Medical Education: A BEME Review

Contents

Team members

Sara Mortaz Hejri, MD, PhD, AFAMEE, Assistant Professor at Education Development Center, Department of Medical Education, Tehran University of Medical Sciences, Tehran, Iran (**Lead author**)

Her major area of interest is clinical competence assessment and most of her academic products have focused on this issue. She teaches workplace based assessment in different faculty development courses. She is a member of TUMS BICC (BEME International Collaborating Center) since 2014. She has also served on the BEME Review Committee (BREC) since 2015. She also had contribution in two Cochrane reviews.

smortaz@tums.ac.ir

Mohammad Jalili, MD, Professor, Department of Emergency Medicine, Department of Medical Education (second affiliation), Vice Chancellor for Education in Tehran University of Medical Sciences, Tehran, Iran

He is a member of TUMS BICC since 2014. While emergency physician by training, he has interest and studies in medical education. His field of interest is assessment and his research activities are mainly focused on this area. He also has experience with systematic review and has already contributed to three published systematic reviews (although none has been in medical education).

mjalili@tums.ac.ir

Mandana Shirazi, PhD, Associate Professor, Education Development Center, Department of Medical Education, Tehran University of Medical Sciences, Tehran, Iran; Affiliated Associate Professor of Department of Clinical Science and Education at SOS Hospital, Karolina Institute, Stockholm, Sweden

She has been teaching in BEME courses in TUMS more than eight years. She has been supervising two thesis which are systematic reviews. She is a member of TUMS BICC since 2014.

Mandana.Shirazi@ki.se

Rasoul Masoomi, PhD candidate in Medical Education, Department of Medical Education, Tehran University of Medical Sciences, Tehran, Iran

He has a bachelor's degree in medical library and information science and has contributed in two systematic reviews. He is also a member of TUMS BICC since 2014.

r-masoomi@razi.tums.ac.ir

Saharnaz Nedjat, Professor, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

She is a member of TUMS BICC and has many medical education papers. She has conducted several systematic reviews; two of them are Cochrane reviews. She has also published many articles on knowledge translation.

nejatsan@tums.ac.ir

John Norcini, PhD, President and Chief Executive Officer of Foundation for Advancement of International Medical Education and Research (FAIMER), Philadelphia, Pennsylvania, USA. Before joining FAIMER, he spent 25 years with the American Board of Internal Medicine where the mini-CEX was proposed for the first time.

JNorcini@faimer.org

Address: No. 57, Hojjatdoust St., Keshavarz Blvd., Tehran, Iran

Telephone: +98-21-88955846

Sources of support: Tehran University of Medical Sciences

Abstract

Background

One of the most frequently-used assessment tools that measure the trainees' performance in workplace is the mini-Clinical Evaluation Exercise (mini-CEX), in which an expert observes and rates the actual performance of trainees.

Several primary studies have evaluated the effectiveness of mini-CEX by assessing its educational and psychometric properties. There are also a number of systematic reviews that have evaluated some aspects of this assessment tool, but none of them adopted a comprehensive framework in order to investigate all characteristics at the same time. A common framework for this purpose is the utility formula, composed five criteria, namely, validity, reliability, educational impact, acceptability and cost.

Objective

The objective of this BME review is to explore, analyze and synthesize the evidence considering the utility of the mini-CEX for assessing undergraduate and postgraduate medical trainees.

Methods

Studies reporting on Mini-CEX performed in undergraduate and postgraduate medical education and providing some empirical data for Mini-CEX in relation to one or more of the validity, reliability, educational impact, acceptability, and cost of mini-CEX, will be included in the review. No restrictions on study design or publication date or language will be handled. To ensure comprehensiveness of our search, we will use different approaches and methods. In addition to electronic search in bibliographic databases, we will conduct forward and backward searching. We will also contact leading authors in the field of Mini-CEX and will search for the grey literature. Data extractions will be done independently by two coders based on a form. If there is any discordance, a third author will resolve it.

The quality assessment will be also done independently by two team members, based on critical appraisal checklists. In attempting to answer our original research questions, we will use meta-analysis or meta-synthesis.

Background

Role of assessment

Assessment plays a central role in medical curriculum. It completes learning process by measuring students' progress and achievement regarding the curriculum outcomes. Several tools have been developed for serving this purpose. Some of these methods focus on cognitive domain of learning and require students to present their knowledge in basic or clinical sciences in written or oral exams. While new assessment methods emphasizes on assessing students' clinical skills, either in a simulated setting dealing with standardized patients, or in a workplace setting encountering the actual patients.

Mini-CEX

One of the most frequently-used assessment tools that measure the trainees' performance in workplace is the mini-Clinical Evaluation Exercise (mini-CEX). In its original form, the mini-CEX is a 9-point rating scale organized in three levels of unsatisfactory (1-3), satisfactory (4-6) and high satisfactory (7-9). An expert, usually a faculty member, observes the actual performance of trainees, rates their history taking and physical examination skills, and provides feedback to them (Norcini et al. 1995). Often it is required that different experts rate several clinical encounters of a trainee throughout the course, rather than one single occasion to be observed by one individual rater.

Following development of the mini-CEX by the American Board of Internal Medicine (ABIM) in the 1990s, it has been widely used in undergraduate and postgraduate medical education programs around the world (Norcini & Burch 2007), both for formative and summative purposes (Weston & Smith 2014).

Scoping review of the primary research on mini-CEX

Several primary studies have evaluated the effectiveness of mini-CEX by assessing its educational and psychometric properties.

- Durning et al. evaluated the reliability and validity of the mini-CEX, which they called mCEX in their paper. Twenty three first-year residents were rated by 46 faculty members. In order to determine validity of the tool, the mini-CEX scores were compared with scores from corresponding sections of a modified version of the ABIM monthly evaluation form (MEF), which demonstrated strong correlations. Reliability measured by calculating coefficient alpha, was found to be 0.90. This study suggested that the mini-CEX was a valid and reliable evaluation tool (Durning et al. 2002).

- Kogan et al. performed a research to know if mini-CEX is feasible, reliable and valid for evaluating clinical skills of medical students in core clerkships which were organized into four 12-week blocks in both inpatient and outpatient settings. In this study, each student was required to complete nine mini-CEX from faculty members and residents. To determine feasibility, the percentage of completed forms, completion times and satisfaction ratings were calculated. To determine reliability,

Cronbach's alpha was calculated and a decision-study also followed a generalizability study. For validity assessment, the correlation between mean scores of mini-CEX, with exam scores and final course grades were calculated. According to the results, 162 students were included in the study, for each of them an average of 8 forms were completed. The reliability coefficient for eight forms was 0.77. Mean mini-CEX scores were significantly correlated with exam scores and final course grades (Kogan et al. 2003).

- Holmboe et al. studied the construct validity of the mini-CEX which was defined as the ability of the mini-CEX to differentiate between levels of performance. Videotapes that portrayed three levels of performance for three clinical skills (history taking, physical examination, and counseling) were developed. Forty faculty members were included from 16 internal medicine residency program. They watched nine videos and evaluated the standard residents using the Mini-CEX form. According to the findings, participants could discriminate well between three levels for the different area of rating; however, range of rating was varied. It seems despite of confirming construct validity of mini CEX as an assessment tool, observation skills of faculty members should be improved in order to decrease inter-rater variability (Holmboe et al. 2003).

- In a qualitative phenomenological study in Buenos Aires, Argentina, Alves de Lima et al. evaluated the educational impact of mini-CEX on the learning of a cohort of cardiology residents. They focused specifically on preparation, regulation, and affective strategies, as well as their appraisal of the exam method. In this study, each of the 16 cardiology residents undertook a one-encounter mini-CEX 45 days after receiving a memorandum with the instructions and was then interviewed (open and semi-structured interview) during the following 72 hours (Alves de Lima et al. 2005).

- Nair et al. carried out an observational study on 209 patient encounters in three teaching hospitals in which the mini-CEX was used for assessing performance of international medical graduates. Generalizability analysis and written surveys were utilized to determine the reliability and acceptability of the mini-CEX, respectively. Learners were satisfied of feedback provided to them and raters believed that feedback is the crucial part of the tool in their workplace setting. The Mini CEX was found to be a reliable and feasible tool for performance assessment of international medical graduates in workplace settings which could provide frequent observations and feedback and also could be applied for summative assessment of learners. (Nair et al. 2008).

- Sidhu et al. carried out a research to determine the reliability and acceptability of the mini-CEX. 188 clinical encounters of fifteen practicing physicians were rated in their office by six previously-trained raters using a validated mini-CEX form. The generalizability coefficient for 10 encounters was 0.92. Ninety four percent of the raters and 75% of physicians expressed in the survey that the mini-CEX is an acceptable assessment (Sidhu et al. 2009).

- Ney et al. who used the abbreviation of mCEX, have attempted to evaluate the predictive validity of mini-CEX by calculating the correlation of obtained scores with future exams. They worked on data from students attending the eight-week internal medicine clerkship in 2006 and 2007. Each student was observed and rated by four house-staffs and four faculty members according to the original mini-CEX form. Since four encounters were enough to fulfill the minimal competency, the authors decided to include students with at least four encounters. Actually, 244 students' scores on mini-CEX were compared with their performance in two standardized patient (SP) exams: end-of-clerkship exam, and end-of-year exam. The findings of this study demonstrate that reliability coefficient for the mini-CEX was 0.90. The correlation between overall mini-CEX scores and end-of-clerkship and end-of-year SP exams were 0.15 (P value=0.018) and 0.21 (P value=0.001), respectively. The authors conclude that performance on the mini-CEX has a small, but modest, correlation with overall future clinical performance in short term and longer term (Ney et al. 2009).

- Brazil et al conducted a study in which an intern cohort was evaluated during their rotation in the emergency department at around mid-term using four sets of mini-CEX. The same interns were also evaluated using Resident Medical Officer (RMO) Assessment Form as their standard in-training assessment. The investigators compared the results of the two assessment methods and also recorded the time taken to perform the mini-CEX assessments as well as the interns' and the assessors' satisfaction and their perception of the feasibility and educational impact of the Mini-CEX. Acceptability was assessed using interns' and assessors' satisfaction as well as their perception of acceptability. The educational impact of the assessment process was enquired from both assessors and interns. Furthermore, the value of adding mini-CEX assessments to the existing assessment processes in summative evaluation was measured considering the number of interns who would have identified as underperforming and would have failed the rotation through the addition of the mini-CEX. Feasibility of mini-CEX was judged according to the time taken for the observation and feedback, as well as the assessors' and interns' perception of feasibility including practical difficulties in arranging and conducting it. The rate of failure to undertake the Mini-CEX according to the schedule was also recorded. The additional direct cost of adding mini-CEX to interns' performance assessment was calculated, too. The authors concluded that adoption of mini-CEX may depend on the availability of resources, the value placed upon the formative purpose, and the acceptability to learners and assessors. They also suggested that feasibility and acceptability of mini-CEX requires further study (Brazil et al. 2012).

- Goel and Tejinder conducted a study to discover if mini CEX was feasible and acceptable, and whether it was helpful in improving learning. For achieving this purpose, pediatric residents were frequently evaluated through a year in different settings of outpatients, wards, and neonatal intensive care units, and were given feedback regarding each encounter. At the end of the year, the opinion of residents and raters was asked using separate forms. The findings showed that mini-CEX did

not induce any unusual stress and overall acceptance was good. Participants found mini-CEX feasible in most clinical settings. The residents found the feedbacks helpful for their learning improvement (Goel and Tejinder, 2015).

Lessons learned from scoping review

Conducting the scoping review helped us learn a couple of lessons which were useful for developing this protocol and performing the review in future.

- We realized that different studies have used various keywords for the instrument, including mini-CEX, mCEX, DOCS, mini-Clinical Exam, etc.

- We also noticed that the studies have targeted different populations (undergraduate, post-graduates, practicing physician, and so on) and have been conducted in different contexts (pediatrics, emergency medicine, etc.).

- The studies also varied in several other aspects: Either original or modified version of the mini-CEX form has been used in different studies. The tool has been used for different purposes (formative vs. summative assessment). Different number of encounters has been considered adequate. Raters have been different (faculty members, senior residents, etc.). The length of rotation in which mini-CEX was used varied. Various number of forms have been filled for each learner. Finally, different outcomes have been evaluated.

We used these lessons in developing our search strategy and also incorporated these findings in our data extraction form.

Previous systematic reviews

A number of systematic reviews have tackled the issue of workplace based assessment (WPBA) tools (appendix 1). Four of them, which we will elaborate on below, included mini-CEX besides other WPBA tools:

- Kogan and colleagues published a systematic review in which they tried to identify assessment tools used for direct observation of clinical skills in settings with actual patients. They excluded the instruments used to assess surgical or procedural skills. In their attempt to evaluate the tools, the reviewers looked for validity evidence of assessment tools using the unitary theory of Messick for construct validity. Hence, they considered the content, concurrent, and predictive validity of the tools, as well as the response process, reliability (in terms of internal consistency, test-retest reliability, inter-rater reliability, and generalizability), and the educational outcomes. In order to assess the educational impact of the instruments participation, self-assessed modification of learner or observer's attitudes, knowledge, or skills, in addition to transfer of learning, and change in organizational delivery or quality of patient care were considered. Furthermore, cost of the tool development and implementation was taken into consideration (Kogan et al. 2009).

- Another systematic review in this field, conducted by Miller and Archer, investigated the literature for evidence of WPBA tools' effect on doctors' performance. They included Mini-CEX, Direct Observation of Procedural Skills (DOPS), Case Based Discussion (CBD), and Multi-Source Feedback (MSF) in their study and focused on postgraduate medical students (residents) and excluded studies on non-medical doctors or medical students. They used Barr's adaptation of Kirkpatrick's four-level model to evaluate the educational impact of WPBA. From 16 studies included, four studies had examined the mini-CEX (Miller and Archer 2010).

- Moreover, Pelgrim et al. reviewed WPBA tools used by professionals to assess directly observed performance of undergraduate or postgraduate medical trainees during authentic clinical encounters, and reported instruments' characteristics as well as their feasibility, reliability, validity, and educational effect. Pilgrim's study did not focus specifically on Mini-CEX and included all similar instruments altogether (Pelgrim et al. 2010).

- The report published by Mills et al. provided evidence on the utility of practice-based assessments in pharmacy pre-registration training. The analytical framework used to review and synthesize evidence was the van der Vleuten utility index (Mills et al. 2011).

Furthermore, there are three reviews that have focused solely on mini-CEX:

- One is a literature review by Hawkins et al. which analyzed the mini-CEX scores within a validity argument framework in terms of scoring, generalization, extrapolation, and interpretation/decision. The search was conducted in MEDLINE only and the inclusion criteria was investigations that had used all or some of the items on the mini-CEX scale without significant changes in scale structure or item descriptors. So, instruments that were significantly modified from the original format had been excluded. The authors stated that except for the area of scoring, other components of the argument generally appeared to be supportive (Hawkins et al. 2010).

- A meta-analysis by Ansari et al. aimed at determining the construct and criterion validity of the mini-CEX. The reviewers evaluated the use of the mini-CEX to assess medical students' or residents' clinical skills in comparison with those participants' use of other clinical measures at various training levels. They included studies that had used the original seven-item version of the mini-CEX and had been published in a peer-reviewed journal. Papers that had focused on generalizability analysis or investigation of the internal structure of the mini-CEX were excluded from the meta-analysis (Ansari et al. 2013).

- Another systematic review conducted by Sandilands and Zumbo that discusses validity theory and validation practices. They restricted their search to the literature published in English language. Articles whose main purposes were to investigate

other assessment tools but also mentioned Mini-CEX were excluded from the review. The reviewers found 13 studies that had investigated the validity of the mini-CEX. However, the authors emphasized that, by conducting this review, they wanted to provide an example to show the (mis)alignment of perspective in current research with the contemporary validity theory, rather than evaluating the quality of the Mini-CEX (Sandilands and Zumbo 2014).

As noted, each of the above-mentioned systematic reviews focused on some characteristics of the Mini-CEX. Moreover, none of the studies adopted a comprehensive framework to analyze its utility.

Utility

A common framework used to explore "all" aspects of assessment tools is the utility formula proposed by van der Vleuten in 1996. Several individual studies and systematic reviews have applied this framework to discuss the effectiveness of different assessment tools (Mills et al. 2011, Ahmed et al. 2011, Naeem et al. 2013).

The framework is composed of validity, reliability, educational impact, acceptability and the cost of the assessment tool (van der Vleuten 1996):

- Reliability refers to the reproducibility and consistency of test results, which might be either the scores obtained from an assessment or pass/fail decisions, for repeating assessment in different conditions. Reliability may have been influenced by different sources of error. So, it can be evaluated by different methods: the agreement between two assessors (inter-rater reliability), the correlation of test items (internal consistency), and the correlation of results when applied at two different time points (test-retest reliability). Moreover, generalizability coefficient has been calculated and reported in recent studies which consider all sources of error at the same time. The concept of reliability is generally expressed as a coefficient ranging from 0 (not reliable) to 1 (perfectly reliable).

- Validity provides rationales and evidence to support the adequacy and appropriateness of interpretation of assessment results. Just as reliability, validity can be evaluated in different ways. However, since the mini-CEX occurs in workplace, we think face validity which determine if the test resembles the situation in the real world, would not be an outcome of interest. So, the emphasis would be on content validity (extent to which the purported content domain is being actually measured), construct validity (extent to which the test is able to differentiate between different groups of performers, and concurrent and predictive validity (correlation of results with gold standard tests and learners future performance, respectively).

- Educational impact, which is also known as consequential validity, demonstrates the effort to recognize the role of assessment in driving students' learning. It specifies that an assessment tool should develop a context aim at improving students' learning. The

assessment scores and also the feedbacks given during the assessment should stimulate change in learners' knowledge, skill, attitude, or performance.

- Acceptability is concerned about to which extent the learners, faculty members, and institutions accept the assessment method which are dealing with. If the test does not come to be accepted by stakeholders, is unlikely to be used successfully and last long.

- Cost involved direct and indirect requirements for implementation of an assessment tool. It considers issues such as the number of raters, training of raters, faculty time, administrative equipment, or IT development.

As can be seen, this framework provided useful and comprehensive criteria beyond traditional psychometrics of validity and reliability. Hence, in this BEME review, we are going to evaluate the Mini-CEX as a method of assessment in the undergraduate and postgraduate medical education, using van der Vleuten formula for utility in order to investigate all above-mentioned characteristics.

Review questions, objectives and keywords

The main objective of this BME review is to explore, analyze and synthesize the evidence considering the utility of the Mini-CEX for assessing undergraduate and postgraduate medical trainees.

The specific objectives are to:

- explore, analyze and synthesize the evidence considering the "validity" of the Mini-CEX for assessing undergraduate and postgraduate medical trainees

- explore, analyze and synthesize the evidence considering the "reliability" of the Mini-CEX for assessing undergraduate and postgraduate medical trainees

- explore, analyze and synthesize the evidence considering the "educational impact" of the Mini-CEX for assessing undergraduate and postgraduate medical trainees

- explore, analyze and synthesize the evidence considering the "cost" of the Mini-CEX for assessing undergraduate and postgraduate medical trainees

- explore, analyze and synthesize the evidence considering the "acceptability" of the Mini-CEX for assessing undergraduate and postgraduate medical trainees

Hence, the following research questions have been developed:

- What is the validity evidence for Mini-CEX in the assessment of undergraduate and postgraduate medical trainees?

- What is the reliability of Mini-CEX in the assessment of undergraduate and postgraduate medical trainees?

- What is the educational impact of Mini-CEX on the undergraduate and postgraduate medical trainees?

- What is the cost of using Mini-CEX in the assessment of undergraduate and postgraduate medical trainees?

- How acceptable is Mini-CEX to medical students and faculty members in undergraduate and postgraduate settings?

Considering the results of our scoping search, we have identified the following keywords for this BEME review:

Population:

- Undergraduate medical trainee, Undergraduate medical education, basic medical education, Medical student
- Postgraduate medical trainees, graduate medical education, residency training, resident

Activity:

- mini-Clinical Evaluation Exercise, Mini-CEX, mCEX, Direct Observation of Clinical Skills, DOCS, Clinical Evaluation Exercise, CEX

Outcome:

- Utility
- Validity, credibility
- Reliability, generalizability, reproducibility, consistency, accuracy
- Educational impact, educational effect, learning impact, educational outcome, consecutive validity
- Cost, cost-effectiveness, feasibility
- Acceptability

For the purpose of this study, the above-mentioned keywords would be defined as follows:

- Undergraduate medical trainees are students undertaking undergraduate or basic medical education at a medical school in order to reach a primary qualification in medicine.

- Postgraduate medical trainees are learners of educational programs for medical graduates entering a specialty. They include formal specialty training as well as academic work in the clinical sciences.

- Mini-CEX or mini-Clinical Evaluation Exercise is an assessment tools used by supervisors in workplace settings to assess clinical performance and provide feedback on a direct observation basis.

- Utility is composed of five components including validity, reliability, educational impact, acceptability, and cost. These terms have been described in background section in details (pages 9 & 10).

Study selection criteria

The inclusion/exclusion criteria have been summarized in table 1.

Studies reporting on Mini-CEX performed in undergraduate and postgraduate medical education and providing some empirical data for Mini-CEX in relation to one or more of the validity, reliability, educational impact, acceptability, and cost of mini-CEX, will be included in the review.

In addition to the original forms of the Mini-CEX, modified versions will also be included in the review, but a subgroup analysis would be done, if needed.

The studies reported multiple data regarding a variety of WPBA tools, only included if Mini-CEX data would have been reported separately.

No restrictions on study design or publication date or language will be handled.

The following studies will be excluded:

- Studies on trainees of disciplines other than medicine
- Studies on Continuous Medical Education/Continuing Professional Development (CME/CPD)
- Studies describing non primary empirical research
- Studies on other WPBA tools such as:
    - Objective Structured Clinical Examination (OSCE)
    - Direct Observation of Procedural Skills (DOPS)
    - Multi-Source Feedback (MSF)
    - Case Based Discussion (CbD)

Table 1: Inclusion and exclusion criteria

	Inclusion Criteria	Exclusion criteria
Population	Studies on undergraduate or postgraduate medical trainees	Studies on non-medical students Studies on CME/CPD

Activity	Studies on the Mini-CEX	Studies on OSCE, DOPS, MSF, or CbD
Outcome	Studies providing data on validity, reliability, educational impact, acceptability, or cost.	
Study Language	All languages	No limitation
Study type	All designs	Papers describing non primary empirical research (Letters and editorial papers)

Search sources and strategies

To ensure comprehensiveness of our search, we will use different approaches and methods. In addition to electronic search in bibliographic databases, we will conduct forward and backward searching by checking the reference lists and citations of the included articles for additional relevant studies. We will also contact leading authors in the field of Mini-CEX. Moreover, in order to find the grey literature, we will search ProQuest Dissertations & Theses and OpenGrey.

Regarding electronic search, the following bibliographic databases will be explored:

- MEDLINE/PubMed
- EMBASE
- CINAHL
- ERIC
- PsycINFO
- SCOPUS
- Web of Sciences

We decided to conduct a scoping search in Ovid to help us make decisions on the final search strategy. While a search strategy is generally supposed to include the population, activity and outcomes, we performed our scoping search only by including mini-CEX and its related terms, because we presumed that the number of primary studies on the mini-CEX would not be substantial. The results have been presented in table 2.

Table 2: Scoping Search for Medline (via Ovid)

	Terms	**number of retrieved titles**
# 1	mini-CEX.mp.	120
# 2	mCEX.mp.	7
# 3	miniCEX.mp.	4
# 4	mini-clinical evaluation exercise*.mp.	92
# 5	miniclinical evaluation exercise*.mp.	3
# 6	clinical evaluation exercise*.mp.	124
# 7	mini clinical exam*.mp.	14
# 8	direct observation of clinical skill*.mp.	8
# 9	workplace based assessment*.mp.	228
# 10	work based assessment*.mp.	52
# 11	work place based assessment*.mp.	6
# 12	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11	**426**

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present
Search Date: 30 September 2016.

As can be noted, less than 500 studies were found. So we assumed that including population (i.e. undergraduate and postgraduate medical trainees), together with study outcomes (reliability, validity, etc.), would diminish the sensitivity of search strategy and limit the number of retrieved articles even further.

Procedure for extracting data

In order to extract and present data from the primary studies, we have designed a data extraction form as can be seen in table 3:

Table 3: The primary version of the data extraction form

Paper code	first author name	Publication date	Journal	Study design	country	postgraduate or undergraduate trainees	Sample size	Min-CEX characteristics	Frequency of use	Summative or formative purposes	Validity	reliability	educational impact	acceptability	cost	Study quality
001																
002																
003																
004																

The extraction form might be revised after checking the selected articles. Before starting the extraction, we will check the reliability of this form. So, two authors independently review 10 articles using this form, and the Kappa coefficient will be calculated for each item. The items with Kappa less than 0.7 will be modified or rechecked. Then, the final form will be approved.

All the extractions will be done independently by two coders based on the final form. If there is any discordance, since it is possible that the source of the disagreement is an error by one side, in the first place, the coders will be asked to discuss the issue. However, if the disagreement cannot be resolved at this stage, a third reviewer will be consulted. He/she will independently extract data, and then discuss with two coders to reach consensus. If consensus cannot be reached, the study authors' will be contacted for further information, and the final decision will be made.

At the end, the consistency between coders will be checked by calculating Kappa coefficient for each item.

Study quality assessment

To evaluate the methodological quality of quantitative studies, we will use the BEME quality framework which consists of 11 indicators (Buckley et al. 2009). Each indicator will be rated as "met," "unmet," or "unclear." In order to be deemed of high quality, studies should meet a minimum of seven indicators (table 4).

Table 4: BEME Quality Indicators (Buckley et al, 2009)

No	Category	Question
1.	Research Question	Is the research question or hypothesis clearly stated?
2.	Study Subjects	Is the subject group appropriate for the study being carried out?
3.	Data Collection Methods	Are the methods used appropriate for the research question and context?
4.	Completeness of data	Attrition rates/acceptable questionnaire response rates
5.	Risk of bias assessment	Is a statement of author positionaity and a risk of bias assessment included?
6.	Analysis of results	Are the statistical and other methods of results analysis used appropriate?
7.	Conclusions	Is it clear that the data justify the conclusions drawn?
8.	Reproducibility	Could the study be repeated by other researchers?
9.	Prospective	Is the study prospective?
10.	Ethical Issues	Are all ethical issues articulated and managed appropriately?
11.	Triangulation	Were results supported by data from more than one source?

To evaluate the methodological quality of qualitative studies, the Critical Appraisal Skills Program (CASP) checklist for the reporting of all qualitative studies will be applied (table 5). The coder will be asked to record a "Yes", "No", or "Can't tell" response.

Table 5: the Critical Appraisal Skills Program (CASP) checklist for quality assessment of qualitative studies

No	Question
1.	Was there a clear statement of the aims of the research?
2.	Is a qualitative methodology appropriate?
3.	Was the research design appropriate to address the aims of the research?
4.	Was the recruitment strategy appropriate to the aims of the research?
5.	Was the data collected in a way that addressed the research issue?
6.	Has the relationship between researcher participants been adequately considered?
7.	Have ethical issues been taken into consideration
8.	Was the data analysis sufficiently rigorous?
9.	Is there a clear statement of findings?
10.	How valuable is the research

These checklists will guide the team members for deciding about the quality of the articles and might be completed or tailored according to the included studies' type.

The quality assessment will be done independently by two team members. Disagreements, similar to the aforementioned procedure, will be resolved via discussion between two coders, and, if required, a third reviewer's opinion will be sought.

We will not remove any study due to weak methodology, though we will consider the quality for quantitative subgroup analysis (if the meta-analysis is possible) or narrative synthesis.

Synthesis of extracted evidence

First, we will provide a description on the characteristics, setting and context of the included studies. This descriptive synthesis will be used as the basis of synthesis evidence in order to address the review questions and objectives.

In attempting to answer our original research questions, we will synthesize the findings to discuss the utility of Mini-CEX in undergraduate and postgraduate medical training. We will present and discuss our findings according to five outcomes (validity, reliability, acceptability, educational impact, and cost).

According to the definitions provided in pages 9 and 10, it is expected to find quantitative data mostly for questions concerning reliability and cost of the mini-CEX. Similarly, it is supposed to have both qualitative and quantitative data for questions regarding validity, educational impact and acceptability.

Considering quantitative data, we predict that there would be significant heterogeneity among the studies that preclude meta-analysis. It means that due to variation in studies' setting, design, and methodology meta-analysis would not be appropriate. However, we will test the statistical heterogeneity by I^2 and X^2 and check it visually through Forest plot. If the findings can be quantitatively synthesized, the random effects model and subgroup analysis (regarding the study quality and study population) will be applied. Otherwise, if there is heterogeneity, we will report the study findings narratively and will undertake a rich and exploratory descriptive synthesis of evidence to explain differences in findings.

Regarding qualitative data, we will use meta-synthesis. Meta-synthesis is an evolving methodology developed to enable systematic synthesis of qualitative data which allows researchers to make explicit the layers of interpretation. We will explore iterative themes within the data and deductively address some of the research questions. As mentioned earlier, a number of our research questions and objectives will be addressed by this approach.

Translation into practice

The findings of this study can be transferred to the medical education stakeholders such as administrators of medical schools, residency program directors and faculty members through brief reports (1:3:25 format) or discussion meetings.
Direct observation of medical trainees who deal with actual patients in real workplace is important for performance-based assessment. We are going to evaluate the utility of Mini-CEX in the light of the existing literature through a systematic approach. If the mini-CEX proves to be useful in evaluating the performance of medical trainees, it can be recommended as part of the formative or even summative assessment in educational programs.
We also hope that publication of this review will encourage stakeholders who have already adopted the mini-CEX to evaluate and report its different characteristics. Lastly, we expect that by conducting a thorough analysis of the psychometric properties of this instrument, we can identify gap of knowledge in this field and suggest areas for future research.

Project timetable

The study will be conducted according to the Gantt chart shown in table 6.

Table 6: project timetable

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Searching	*	*	*	*														
Data extraction & coding					*	*	*	*	*	*								
Data synthesis											*	*	*	*				
Preparing the manuscript															*	*	*	*

Conflict of interest statement

The authors report no conflict of interest.

References

1. Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB. Observational tools for assessment of procedural skills: a systematic review. Am J Surg. 2011;202(4):469-480.
2. Alves de Lima A, Henquin R, Thierer J, Paulin J, Lamari S, Belcastro F, Van der Vleuten CP. A qualitative study of the impact on learning of the mini clinical evaluation exercise in postgraduate training. Med Teach. 2005 ;27(1):46-52
3. Ansari A, Ali SK, Donnon T. The construct and criterion validity of the mini-CEX: a meta-analysis of the published research. Acad Med 2013;88(3):413-20.
4. Brazil V, Ratcliffe L, Zhang J, Davin L. Mini-CEX as a workplace-based assessment tool for interns in an emergency department-does cost outweigh value? Med Teach. 2012;34(12):1017-23.
5. Buckley S, Coleman J, Davison I, Khan KS, Zamora J, Malick S, et al. The educational effects of portfolios on undergraduate student learning: a Best Evidence Medical Education (BEME) systematic review. BEME guide no. 11. Med Teach. 2009;31(4):282–98.
6. Durning SJ, Cation LJ, Markert RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. Acad Med. 2002;77(9):900-4.
7. Goel A, Singh T. The usefulness of Mini Clinical Evaluation Exercise as a learning tool in different pediatric clinical settings. Int J Appl Basic Med Res. 2015;5(Suppl 1:S32-4.
8. Hawkins RE, Margolis MJ, Durning SJ, Norcini JJ. Constructing a validity argument for the mini-Clinical Evaluation Exercise: a review of the research. Acad Med 2010;85(9):1453-61
9. Holmboe ES, Huot S, Chung J, Norcini J, Hawkins RE. Construct validity of the miniclinical evaluation exercise (miniCEX). Acad Med. 2003;78(8):826-30.
10. Jelovsek JE, Kow N, Diwadkar GB. Tools for the direct observation andassessment of psychomotor skills in medical trainees: a systematic review. Med Educ 2013;47(7):650-73.
11. 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.
12. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment ofclinical skills of medical trainees: a systematic review. JAMA 2009;23;302(12):1316-26
13. Miller A, Archer J. Impact of workplace based assessment on doctors' educationand performance: a systematic review. BMJ 2010;24;341:c5064.
14. Mills E, Blenkinsopp A, McKinley RK, Black P. The Assessment of Observed Practice: A Literature Review. University of Keele2011

15. Morris MC, Gallagher TK, Ridgway PF. Tools used to assess medical students competence in procedural skills at the end of a primary medical degree: a systematic review. Medical Education Online 2012;17:10.3402

16. Naeem N. Validity, reliability, feasibility, acceptability and educational impact of direct observation of procedural skills (DOPS). J Coll Physicians Surg Pak. 2013;23(1):77-82.

17. Nair BR, Alexander HG, McGrath BP, Parvathy MS, Kilsby EC, Wenzel J, Frank IB, Pachev GS, Page GG. The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates. Med J Aust. 2008;189(3):159-61

18. Ney EM, Shea JA, Kogan JR. Predictive validity of the mini-Clinical Evaluation Exercise (mcex): do medical students' mCEX ratings correlate with future clinical exam performance? Acad Med 2009;84(10 Suppl):S17-20

19. Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. Ann Intern Med 1995;123(10):795-9

20. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. Med Teach 2007;29(9):855-71.

21. Pelgrim EAM, Kramer AWM, Mokkink HGA, van den Elsen L, Grol RPTM, van der Vleuten CPM. In-training assessment using direct observation of single-patient encounters: a literature review. Advances in Health Sciences Education. 2011;16(1):131-142.

22. Sandilands DD, Zumbo BD. (Mis)Alignment of medical education validation research with contemporary validity theory: the mini-CEX as an example. In: Validity and validation in social, behavioral, and health sciences. Springer 2014:289–310.

23. Sidhu RS, Hatala R, Barron S, Broudo, M, Pachev G, Page G. Reliability and acceptance of the mini-clinical evaluation exercise as a performance assessment of practicing physicians. Academic Medicine 2009; Journal of the Association of American Medical Colleges, 84(10):S113–S115.

24. Van der Vleuten, CPM. The assessment of professional competence: developments, research and practical implications. Advances in Health Sciences Education 1996: 41-67.

25. Weston PS, Smith CA. The use of mini-CEX in UK foundation training six years following its introduction: lessons still to be learned and the benefit of formalteaching regarding its utility. Med Teach 2014;36(2):155-63.

Appendix 1: Review articles regarding Mini-CEX

Study	Inclusion criteria	Outcome	Limitation
Kogan, 2009	• Studies that used tools for direct observation of skills in clinical settings with actual patients for medical students, interns, residents, or fellows • English-language studies • Published between 1965 and 2009	(based on unitary theory of Messick for construct validity) • content • concurrent, predictive validity • Response process • Reliability: o internal consistency o test-retest reliability o agreement (interrater reliability) o generalizability • Educational outcomes o participation: learners' or observers' views on the tool or its implementation o self-assessed modification of learner or observer attitudes, knowledge, or skills o transfer of learning: objectively measured change in learner or observer knowledge or skills o results: change in organizational delivery or quality of patient care *Plus* • cost of tool development and implementation	• exclusion of instruments used to assess surgical or procedural skills
Miller, 2010	• studies that used workplace based assessment tools: Mini-CEX, DOPS, CBD, and MSF	• the educational impact • effect of workplace based assessment on doctors' performance • using Barr's adaptation of Kirkpatrick's four level evaluation model	• Excluded studies on non-medical doctors or medical students
Hawkins et al. 2010	• Investigations that had used all or some of the items on the mini-CEX scale without significant changes in scale structure or item descriptors.	• A validity argument framework in terms of scoring, generalization, extrapolation, and interpretation/decision.	• Exclusion of modified versions
Mills 2011	• practice-based assessments in pharmacy pre-registration training	• reliability • validity • educational impact • acceptability • cost	
Morris, 2012	• Studies that used procedural	• validity: content and face • reliability/IRR/generalizability	• excluded: OSATS and DOPS (post-

	assessment toolsfor medical studentsprior to certification		registration assessment and clinical rather than simulated competence)
Pelgrim, 2012	• workplace based tools used by professionals to assess directly observed performancein authentic patient encounters forpostgraduate or undergraduate medical trainees	• characteristics • feasibility • reliability • validity • educational effect	
Ansari et al. 2013	• Studies that had used the original seven-item version of the mini-CEX and had been published in a peer-reviewed journal. • medical students' or residents	• construct and criterion validity	
Sandilands and Zumbo 2014	• literature published in English language	• validity	• Restricted to English language • Exclusion of articles whose main purpose was to investigate other assessment tools but also mentioned Mini-CEX were excluded