

The Construct and Criterion Validity of the Mini-CEX: A Meta-Analysis of the Published Research

Ahmed Al Ansari, MHPE, MRCSI, Syeda Kauser Ali, PhD, and Tyrone Donnon, PhD

Abstract

Purpose

To conduct a meta-analysis of published studies to determine the construct and criterion validity of the mini-clinical evaluation exercise (mini-CEX) to measure clinical performance.

Method

The authors included all peer-reviewed studies published from 1995 to 2012 that reported the relationship between participants' performance on the mini-CEX and on other standardized academic and clinical performance measures. Moderator variables and performance and standardized exam measures were extracted and reviewed

independently using a standardized coding protocol.

Results

Performance measures from 11 studies were identified. A random-effects model of weighted mean effect size differences (d) resulted in (1) construct validity coefficients for the mini-CEX on the trainees' performance across different residency year levels ranging from $d = 0.25$ (95% confidence intervals [CI]: 0.04–0.46) to $d = 0.50$ (95% CI: 0.31–0.70), and (2) concurrent validity coefficients for the mini-CEX based on personnel ratings ranging from $d = 0.23$ (95% CI: 0.04–0.50) to $d = 0.50$ (95% CI: 0.34–0.65). Also, a

random-effects model of weighted correlation effect size differences (r) resulted in predictive validity coefficients for the mini-CEX on trainees' performance across different standardized measures ranging from $r = 0.26$ (95% CI: 0.16–0.35) to $r = 0.85$ (95% CI: 0.47–0.96).

Conclusions

The construct and criterion validity of the mini-CEX was supported by small to large effect size differences based on measures between trainees' achievement and clinical skills performance, indicating that it is an important instrument for the direct observation of trainees' clinical performance.

The evaluation of the clinical competence of medical students and residents can best be achieved through the use of direct observation. There has been inconsistency in how best to measure and compare performance on clinical skill domains, however. In response to the problems of the longer clinical evaluation exercise, in 1972 the American Board of Internal Medicine proposed the use of the mini-clinical evaluation exercise (mini-CEX) to evaluate residents in the completion of a patient history and physical examination

that results in the demonstration of organized clinical judgments and efficient counseling skills.¹

The mini-CEX is a seven-item, global rating scale that is designed to evaluate medical students' and residents' patient encounters in about 15 to 20 minutes. The mini-CEX is specifically designed to assess the skills that residents require in actual patient encounters and also to reflect the educational requirements that are expected by attending physicians during teaching rounds.^{2,3} As described by Norcini et al,¹ the multiple use of the mini-CEX with trainees allows for a greater variability across different patient encounters that results in a more reliable and valid measure of clinical skill practice and development. It is a performance-based evaluation method that is used to assess selected clinical competencies (e.g., patient interview and physical examination, communication and interpersonal skills) within a medical training context.

Although the mini-CEX continues to be widely used in a broad range of clinical settings, there are concerns about the reliability and validity of this assessment instrument in evaluating

medical students' and residents' clinical performance. In a recent literature review of in-training assessment using direct observation of single-patient encounters, Pelgrim et al⁴ acknowledged the mini-CEX as one of the best-supported instruments but stated that more evidence of construct validity is needed. Specifically, de Lima et al⁵ computed a range of mean scores on the mini-CEX items for a sample of 108 first- to fourth-year cardiology residents showing modest increases across the years of training. Kogan et al⁶ reported several correlation coefficients ($r = 0.17$ – 0.43) with a sample of medical students ($n = 162$) when comparing their mean mini-CEX scores with other written and clinical performance measures. In addition, Hatala et al⁷ computed a range of correlation coefficients from $r = 0.29$ to 0.60 on the mean score for the mini-CEX with residents' measures on certifying internal medicine oral, bedside, and written exams ($n = 162$). In part because the mini-CEX has been used in a variety of contexts and the research focuses on measures across program years or is based on comparisons with other assessment methods, the validity of this clinical evaluation instrument needs further exploration.

Dr. Al Ansari is senior resident, Department of Surgery, Faculty of Medicine, Bahrain Defense Force Hospital, Riffa, Bahrain.

Dr. Ali is assistant professor, Department of Family Medicine, Faculty of Medicine, Aga Khan University, Karachi, Pakistan.

Dr. Donnon is associate professor, Medical Education and Research Unit, Department of Community Health Sciences, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada.

Correspondence should be addressed to Dr. Donnon, Medical Education and Research Unit, G13 Health Medical Research Building, Faculty of Medicine, University of Calgary, 3330 Hospital Dr., NW, Calgary, AB Canada, T2N 4N1; telephone: (403) 210-9682; fax: (403) 210-7507; e-mail: tdonnon@ucalgary.ca.

Acad Med. 2013;88:0–0.

First published online

doi: 10.1097/ACM.0b013e318280a953

The main purpose of our study, therefore, was to conduct an empirical integration of all published data on the use of the mini-CEX to assess medical students' or residents' clinical skills in comparison or contrasted with those participants' use of other clinical measures at various training levels. In the present study, we conduct a meta-analysis on the construct and criterion (predictive or concurrent) validity of the mini-CEX as a function of both summary effect sizes and interpretation of the magnitude of coefficients as well as their confidence intervals (CIs).

Method

Selection of studies

In this study, we followed the *JAMA* guidelines for the reporting of observational studies included in a meta-analysis.⁸ In addition to a MEDLINE (January 1995 to January 2012) search, we also searched the PsychINFO (January 1995 to January 2012), EMBASE (January 1995 to January 2012), and CINAHL (January 1995 to January 2012) databases. In addition, all lists of articles that used or referenced the use of the mini-CEX were reviewed to ensure that all relevant publications were identified.

To be included, a primary study had to meet the following criteria: (1) It used the original seven-item version of the mini-CEX, (2) it reported empirical findings on the use of the mini-CEX related to either medical students' or residents' clinical performance, (3) when applicable, it employed psychometrically sound criterion measures (e.g., standardized instruments, summative in-training evaluations, objectively scored observational ratings), and (4) it was published in a refereed, peer-reviewed journal. The purpose for restricting the search of the articles to refereed journals was to enhance the inclusion of studies that are of high quality. On the other hand, studies were excluded if (1) the focus of the article was restricted to a generalizability analysis or investigation of the internal structure of the mini-CEX,⁹ (2) the review on the use of the mini-CEX did not provide any new empirical data,¹⁰ and (3) the analysis focused on differences related to rater stringency without reporting on actual student performance outcomes.^{11,12}

Data extraction

Our initial literature review and search of the databases yielded 31 peer-reviewed journal articles; 11 studies met the inclusion criteria requirements, and the other 20 articles failed to meet all the relevant inclusion criteria (e.g., review articles without new data, focus on rater stringency or training, emphasis on factor analysis or generalizability of mini-CEX). A coding protocol was developed that included each study's title, author(s)' name(s), year published, source of publication, study design (i.e., construct or criterion validity study), mini-CEX measures reported (i.e., a single item, all item domains, total mean score), student category (i.e., medical students, residents), program specialty, and types of raters (i.e., faculty, residents, consultants). The following moderator variables were coded when available: sex, age, race, ethnicity, and location of medical school or residency program. All 11 articles were coded independently by two of us (A.A. and S.K.), and any discrepancies (e.g., effect size calculations) were reviewed by another (T.D.). On the basis of our iterative reviews and discussions, we were able to achieve 100% agreement on all coded data.

Statistical analysis

The statistical analysis of all effect size calculations was performed using the Comprehensive Meta Analysis software program (version 1.0.23, Biostat Inc., Englewood, New Jersey). Depending on the empirical data reported in each of the primary studies, we used the Pearson product-moment correlation coefficient (r) or mean differences (Cohen d) as the effect size measures. We selected mini-CEX items or total mean score on the mini-CEX measures as the variables and either contrasted between groups (e.g., postgraduate or in-training year) or compared mini-CEX scores with other clinical skill measures (e.g., in-training evaluation report). For 5 (45%) of the 11 studies included, a correlation coefficient between mini-CEX scores and performance on another outcome was provided in the results section (see Table 1). For 7 (64%) of the studies, we calculated d from other reported data (note that from Kogan and colleagues⁶ study we were able to extract both r and d effect size values).

With the combination of results from studies that used different research designs (e.g., different years of residency training, resident versus faculty ratings of medical students) and methods of analysis between groups (e.g., mini-CEX in comparison with in-training evaluation reports, oral exams, bedside assessments), we used a random-effects model in combining the unweighted and weighted effect sizes. Although a fixed-effect model assumes that the summary effect size differences are the same from study to study (e.g., the consistent use of the mini-CEX instrument), the random-effects model calculation reflects a more conservative estimate of the between-study variance of the participants' clinical skills performance measures.¹³ Forest plots with Cochran Q tests for heterogeneity of effect sizes were completed, but the absence of a significant P value for Q may imply low power within studies rather than actual consistency or homogeneity across the studies included in the meta-analysis.^{14,15} Subsequently, a review of the dispersion of the studies in the forest plots was an important visual indicator for evaluating the consistency between studies. The interpretation of the magnitude of the effect size for both mean differences and correlations is based on Cohen's¹⁶ suggestions: $r = 0.10$ to 0.29 and $d = 0.20$ to 0.49 are "small," $r = 0.30$ to 0.49 and $d = 0.50$ to 0.79 are "medium," and $r > 0.50$ and $d > 0.80$ are considered to be "large" effect size differences.

Results

The characteristics of the 11 studies included in the meta-analysis are based on four groups that we identified (see Table 1) that report contrasts between trainees within one year of a residency program (Group A), differences between performance levels within a peer group (Group B), rating differences between faculty/residents (Group C), and comparisons between the mini-CEX and other measures of achievement or performance (Group D). In addition, the reported mini-CEX domain measure (i.e., items 1–7, or the total mean score) and corresponding unweighted effect sizes based on either the contrast or comparison variables are provided in Table 1. The studies included illustrate different approaches to testing the validity of the mini-CEX. Groups A, B,

T1

and C test the construct validity of the items or mean scores of the mini-CEX instrument by showing that medical students or residents at difference levels or by personnel ratings tend to obtain higher clinical skills performance scores, and Group D tests the criterion validity by comparing the mini-CEX with other similar assessments of clinical performance as either a concurrent or predictive measure.

The sample sizes of the studies range from 9 residents¹⁷ to 244 medical students¹⁸ who had been assessed using the mini-CEX with as few⁵ as 2.3 completed forms and as many¹⁷ as 38 completed forms per individual (note that in Holmboe and colleagues¹⁷ study, 38 faculty members each viewed nine videotaped residents' performance at the poor, marginal, and superior levels). In our meta-analysis, we treated medical students and residents equally in that they represent learners at different stages of their training in clinical skills development. Therefore, we are evaluating the performance of these *trainees* as a function of their ability to conduct an appropriate medical interview, physical examination, etc. The training provided to medical students and residents at the bedside or on the ward adheres to a similar teaching and learning process where the performance of these skills reflects the clinical competency expectations for all clinicians in practice. Information on specific demographic characteristics such as students' sex or age was not reported, but level of training and residency program were typically identified. In each study, the unweighted mean effect size difference (Cohen *d*) or Pearson product-moment correlation (*r*/W^M) is provided between the mini-CEX item or mean score and either a contrasting variable (e.g., in-training level) or comparison measure (e.g., in-training evaluation reports).

Construct validity of the mini-CEX

Of the 11 studies that reported data on medical students' or residents' clinical skills performance, 7 (64%) demonstrate results in support of the construct validity of the mini-CEX. As shown in Table 2, we combined 4 of the studies (Group A) to show that, for each of the seven mini-CEX items, a range of effect size differences in performance between a single year of residency training (e.g., change in ratings as a function of

year 1 to year 2, year 2 to year 3, etc.) from $d = 0.25$ (95% CI, 0.04–0.46) in humanistic qualities/professionalism to $d = 0.50$ (95% CI, 0.31–0.70) in overall clinical competence. As illustrated in the forest plot (Chart 1), the combined fixed-effect and random-effects size for the three Group A studies (and seven outcome measures) for the overall clinical competence item were both shown to be a “medium” effect size difference, $d = 0.50$ (95% CI, 0.31–0.70).

When differences between performance level within a peer group (superior/honors, marginal/high pass, poor/pass) were investigated, we found two studies (Group B) that showed a mean difference in clinical performance on three items and a total mean score of the mini-CEX. In particular, Holmboe et al¹⁷ compared ratings on the medical interviewing, physical examination, and counseling skills items to show that there are consistently large effect size differences between mean ratings of poor, marginal, and superior second-year residents. In particular, the ratings on these items ranged from $d = 0.90$ between superior/marginal residents on the medical interviewing item to a mean difference of $d = 4.00$ between superior/poor residents on the physical exam skills item. Correspondingly, Kogan et al⁶ found a range of mean mini-CEX scores that varied from $d = 0.04$ with high-pass/pass-level medical students in an inpatient setting to $d = 1.00$ with honors/pass-level medical students in an outpatient setting.

In Group C, we combined the outcomes from two studies that investigated the mean differences in ratings provided by residents and faculty members on medical students. As shown in Table 2, the ratings of medical students by faculty are consistently more stringent than those of residents across all seven of the mini-CEX items. As illustrated by the forest plot in Chart 1, residents are more lenient in the mean ratings of medical students' performance on the overall clinical competence item, with a combined “medium” random-effects size difference of $d = 0.38$ (95% CI, 0.15–0.62).

Criterion (predictive/concurrent) validity of the mini-CEX

Of the 11 studies included in the meta-analysis, 5 reported data (Table 3) on either medical students' or residents' mini-CEX item or mean score ratings

with some other criterion measure (e.g., in-training evaluation reports, inpatient or outpatient write-ups, examination checklists). Although the mean effect size differences were found to be “medium” across each of the seven items on the mini-CEX, the total mean score resulted in a combined “small” effect size difference, $d = 0.26$ (95% CI, 0.16–0.35). As shown in the forest plot (Chart 1), the combined random-effects size calculation for the overall clinical competence item was “medium,” $d = 0.64$ (95% CI, 0.48–0.77).

Although the Cochran Q test shows significant heterogeneity between the studies included in Groups A, C, and D on the medical students' and residents' mini-CEX ratings, an analysis to determine the potential differences as a result of moderator variables (e.g., sex, year of program) was limited by the information provided across the primary studies included in the meta-analysis. Nevertheless, the studies are weighted by their respective sample sizes, and the random-effects model analysis (with greater 95% CIs) provides a more conservative estimate of the combined effect sizes and with the overall clinical competence item illustrated in the forest plot diagram.

Conclusion

Our study provided four major findings.

1. The mini-CEX has evidence of construct validity when used with residents across the years of a residency program. Residents' performance on the mini-CEX items across one year of residency training showed “small” to “medium” effect size differences, with the effect sizes ranging from $d = 0.25$ (95% CI, 0.04–0.46) to $d = 0.50$ (95% CI, 0.31–0.70). When performance on items across more than one year of residency training was reviewed, however, the unweighted mean effect size differences were found to be even greater (e.g., $d = 3.80$ between superior/poor-performing residents on the counseling skills item).
2. The effect size differences between performance levels within a peer group (superior/honors, marginal/high pass, poor/pass) ranged from $d = 0.43$ (95% CI, 0.23–0.63) in one study on the total mean score of the mini-

Table 1

Characteristics of 11 Mini-Clinical Evaluation Exercise (mini-CEX) Studies From 1995 to 2012 Identified in a Meta-Analysis of Research, to Determine the Construct and Criterion Validity of the mini-CEX as a Measure of Clinical Performance*

Study source and sample size	Group [†]	Contrast [‡]	mini-CEX domain [§]	Effect size difference ($d_{\text{mini-CEX}}$)
• de Lima et al, ⁵ 2007	A	PGY 1/PGY 2	1, 2, 3, 4, 5, 6, 7	0.57, 0.41, 0.20, 0.17, 0.01, 0.24, 0.41
• Sample size = 108 (PGY 1 = 15, PGY 2 = 38, PGY 3 = 44, PGY 4 = 11), mini-CEX total = 253		PGY 1/PGY 3	1, 2, 3, 4, 5, 6, 7	0.53, 0.52, 0.22, 0.56, 0.17, 0.37, 0.71
		PGY 1/PGY 4	1, 2, 3, 4, 5, 6, 7	1.11, 1.23, 0.66, 0.94, 0.83, 0.87, 1.18
		PGY 2/PGY 3	1, 2, 3, 4, 5, 6, 7	0.00, 0.12, 0.01, 0.36, 0.15, 0.15, 0.30
		PGY 2/PGY 4	1, 2, 3, 4, 5, 6, 7	0.51, 0.74, 0.43, 0.72, 0.78, 0.87, 0.75
		PGY 3/PGY 4	1, 2, 3, 4, 5, 6, 7	0.48, 0.61, 0.38, 0.38, 0.53, 0.46, 0.45
• Norcini et al, ² 1995	A	PGY 1/PGY 2	7	0.88
• Sample size = 87 (PGY 1 = 78, PGY 2 = 9), mini-CEX total = 388				
• Norcini et al, ¹ 2003	A	PGY 1 ^{1st/4th Q}	1, 2, 3, 4, 5, 6, 7	0.50, 0.56, 0.35, 0.58, 0.50, 0.64, 0.56
• Sample size = 183 (PGY 1 ^{1st Q} = 86, PGY 1 ^{4th Q} = 97), mini-CEX total = 1,228				
• Wiles et al, ²³ 2007	A	PGY 1/PGY 2	5, 7	0.00, 0.11
• Sample size = 27 [PRHO (PGY 1) = 5, SHO (PGY 2) = 11, SpR (PGY 3) = 11], mini-CEX total = 124				
• Holmboe et al, ¹⁷ 2003		PGY 1/PGY 3	5, 7	0.54, 0.96
		PGY 2/PGY 3	5, 7	0.55, 0.95
• Sample size = 9 (PGY 2: poor = 3, marginal = 3, superior = 3), mini-CEX total = 348	B	Superior/marginal	1, 2, 5	0.90, 2.20, 1.40
		Superior/poor	1, 2, 5	3.20, 4.00, 3.80
		Marginal/poor	1, 2, 5	2.10, 1.80, 2.40
• Kogan et al, ⁶ 2003	C	Res/fac ratings	1, 2, 3, 4, 5, 6, 7	0.35, 0.43, 0.33, 0.09, 0.43, 0.26, 0.26
• Sample size = 162 (medical students), mini-CEX total = 1,137 (residents = 434, faculty = 703), (medical students: pass = 54, high pass = 54, honors = 54), mini-CEX total = 1,297	B	Honors/high pass	MS (IP, OP, WU, FG)	0.12, 0.50, 0.16, 1.00
		Honors/pass	MS (IP, OP, WU, FG)	0.13, 1.00, 0.57, 0.83
		High pass/pass	MS (IP, OP, WU, FG)	0.04, 0.50, 0.42, 0.00
• Torre et al, ²⁴ 2007	C	Res/fac ratings	1, 2, 3, 4, 5, 6, 7	0.40, 0.29, 0.29, 0.36, 0.56, 0.42, 0.50
• Sample size = 177 (medical students), mini-CEX total = 354 (residents = 205, faculty = 149)				

(Continues)

Table 1
(Continued)

Study source and sample size	Group [†]	Comparison ^{**}	mini-CEX domain [§]	Effect size difference (r_{UNWT})
• Boulet et al, ²¹ 2002	D	CSA COM	1, 3, 5, 6	0.88, 0.87, 0.85, 0.85
• Sample size = 10 (international medical graduates), mini-CEX total = 100				
• Durning et al, ²⁵ 2002	D	IPS1, IPS2	2	0.61, 0.70
• Sample size = 23 (PGY 1), mini-CEX total = 162				
• Kogan et al, ⁶ 2003	D	ABIM ITER	1, 2, 3, 4, 7	0.59, 0.60, 0.59, 0.72, 0.79
• Sample size = 162 (medical students), mini-CEX total = 1,137				
• Hatala et al, ⁷ 2006	D	IP, OP, ES, FG, WU	MS	0.43, 0.35, 0.22, 0.19, 0.17
• Sample size = 22 (PGY 4), mini-CEX total = 120				
• Ney et al, ¹⁸ 2009	D	RCPSC IM–Oral	1, 2, 3, 4, 6, 7	0.44, 0.60, 0.43, 0.45, 0.56, 0.59
• Sample size = 244 (medical students), mini-CEX total = 1,464				
		RCPSC IM BS	1, 2, 3, 4, 6, 7	0.43, 0.56, 0.42, 0.39, 0.53, 0.55
		RCPSC IM MCQ	1, 2, 3, 4, 6, 7	0.47, 0.54, 0.29, 0.50, 0.50, 0.59
	D	CSE CL (Bk1,2,3,4)	MS	0.26, 0.45, 0.07, 0.12

*The table illustrates all of the reported effect size data that the authors extracted from each of the 11 studies. For example, in de Lima and colleagues' article, the authors extracted contrasts in clinical performance (as measured by the mini-CEX) between the residents from PGYs 1 through 4. In the meta-analysis of the effect sizes for Group A, however, residents could not be more than one year apart (e.g., contrasting performance data between PGY 1 and 2, or 2 and 3, or 3 and 4 were used), so the data shown in this table from the lines PGY 1/PGY 3, PGY 1/PGY 4, and PGY 2/PGY 4 were not used at all in the authors' analysis, but those data are shown so that readers can see how large some of these effects sizes were when comparing residents that are two or more years apart in the training program. Note that in this table, the rows of reported data associated with each Group (i.e., A, B, C, and D) were used to calculate a combined effect size and 95% confidence interval for each of the seven mini-CEX items or the "mean score on the mini-CEX" shown in Tables 2 and 3. In those latter tables, the authors included the numbers 1 through 7 (indicating the seven items of the mini-CEX) so that readers can go back to the present table and find, for example, the two studies (and four outcomes) that were used to generate the combined effect size for "1. Medical interview skills" for Group A.

[†]Group A reports contrasts between trainees within one year of a residency program, Group B reports differences between performance levels within a peer group, Group C reports rating differences between faculty/residents, and Group D reports comparisons between the mini-CEX and other measures of achievement or performance.

[‡]PGY indicates postgraduate year of residency program; res/fac, residents/faculty.

[§]1 = medical interview skills, 2 = physical examination, 3 = humanistic qualities/professionalism, 4 = clinical judgment, 5 = counseling, 6 = organization/efficiency, 7 = overall clinical competence. MS indicates mean score on mini-CEX.

^{||}The symbol " d_{UNWT} " refers to the unweighted mean effect size difference as defined by Cohen d. The symbol " r_{UNWT} " refers to the association as defined by Pearson product-moment correlation unweighted mean.

**PRHO indicates preregistration house officer; SHO, senior house officer; SpR, specialist registrar. Achievement level in the summative assessments: IP indicates inpatient; OP, outpatient; ES, exam score; FG, final grade; WU, write-ups; BS, bedside score; ABIM, American Board of Internal Medicine; ITER, in-training evaluation report; CSA, clinical skills assessment; CSE CL, clinical skills examination checklist (clerkship blocks); Bk, block; COM, doctor-patient communication composite; IPS1, interpersonal skills for interviewing; IPS2, interpersonal skills for counseling; IM, internal medicine; RCPSC IM, Royal College of Physicians and Surgeons of Canada Internal Medicine.

Table 2

Random-Effects Model (Cohen d) of the Mini-Clinical Evaluation Exercise (mini-CEX) Checklist Domains With Year/Performance Level (Groups A and B)*

mini-CEX domain measure	Included studies (no. of outcomes)	Sample size	Group A. Difference between one PGY level [†]	Included studies (No. of outcomes)	Sample size	Group B. Difference between performance levels [‡]
1. Medical interview skills	2 (4)	373	0.37 (0.11–0.63)	1 (3)	18	1.34 (0.01–2.70)
2. Physical examination skills	2 (4)	373	0.43 (0.22–0.65)	1 (3)	18	1.86 (0.31–3.40)
3. Humanistic qualities/ professionalism	2 (4)	373	0.25 (0.04–0.46)	—	—	—
4. Clinical judgment	2 (4)	373	0.45 (0.24–0.67)	—	—	—
5. Counseling skills	3 (6)	411	0.35 (0.14–0.55)	1 (3)	18	1.70 (0.21–3.20)
6. Organization/efficiency	2 (4)	373	0.42 (0.16–0.68)	—	—	—
7. Overall clinical competence	4 (7)	498	0.50 (0.31–0.70)	—	—	—
Mean score on the mini-CEX	—	—	—	1 (12)	1,296	0.43 (0.23–0.63)

*Based on data from 11 mini-CEX studies from 1995 to 2012 identified in a meta-analysis of the research, to determine the construct and criterion validity of the mini-CEX as a measure of clinical performance.

[†]PGY indicates postgraduate year. Effect sizes were combined from differences reported within one PGY level to the next PGY level (e.g., PGY 1 to PGY 2, PGY 2 to PGY 3, etc.).^{1,2,5,23}

[‡]Effect sizes combined from academically poor/pass to marginal/high-pass to superior/honors level.^{6,17}

CEX up to $d = 1.86$ (95% CI, 0.31–3.40) on the physical examination skills item. Although these results are based on two separate studies included in the meta-analysis, the findings are based on the combination of 3 and 12 outcome measures combined from each of the studies, respectively.

3. The rating differences of medical students on the mini-CEX between personnel (either residents or faculty members) showed “small” effect size differences that ranged from $d = 0.23$

(95% CI, 0.04–0.50) on the clinical judgment item to $d = 0.50$ (95% CI, 0.34–0.65) on the counseling skills item. In particular, these results support other findings that have shown that, in comparison with faculty evaluators, residents tend to be more lenient and score medical students higher on in-training evaluation checklists.^{19,20}

4. The mini-CEX shows evidence of criterion-related validity when compared with other clinical skill

achievement (e.g., certifying oral and written examinations) or performance (e.g., in-training evaluation reports, inpatient or outpatient write-ups) measures. We found “small” to “large” correlation coefficients with combined effect sizes ranging from $r = 0.26$ (95% CI, 0.16–0.35) on the mean score of the mini-CEX to $r = 0.64$ (95% CI, 0.48–0.77) on the overall clinical competence item.

The construct- and criterion-related validity of the mini-CEX are supported by

Table 3

Random-Effects Model (Cohen d/Correlation r) of the mini-CEX Checklist Domains With Personnel Ratings/Academic Performance (Groups C and D)*

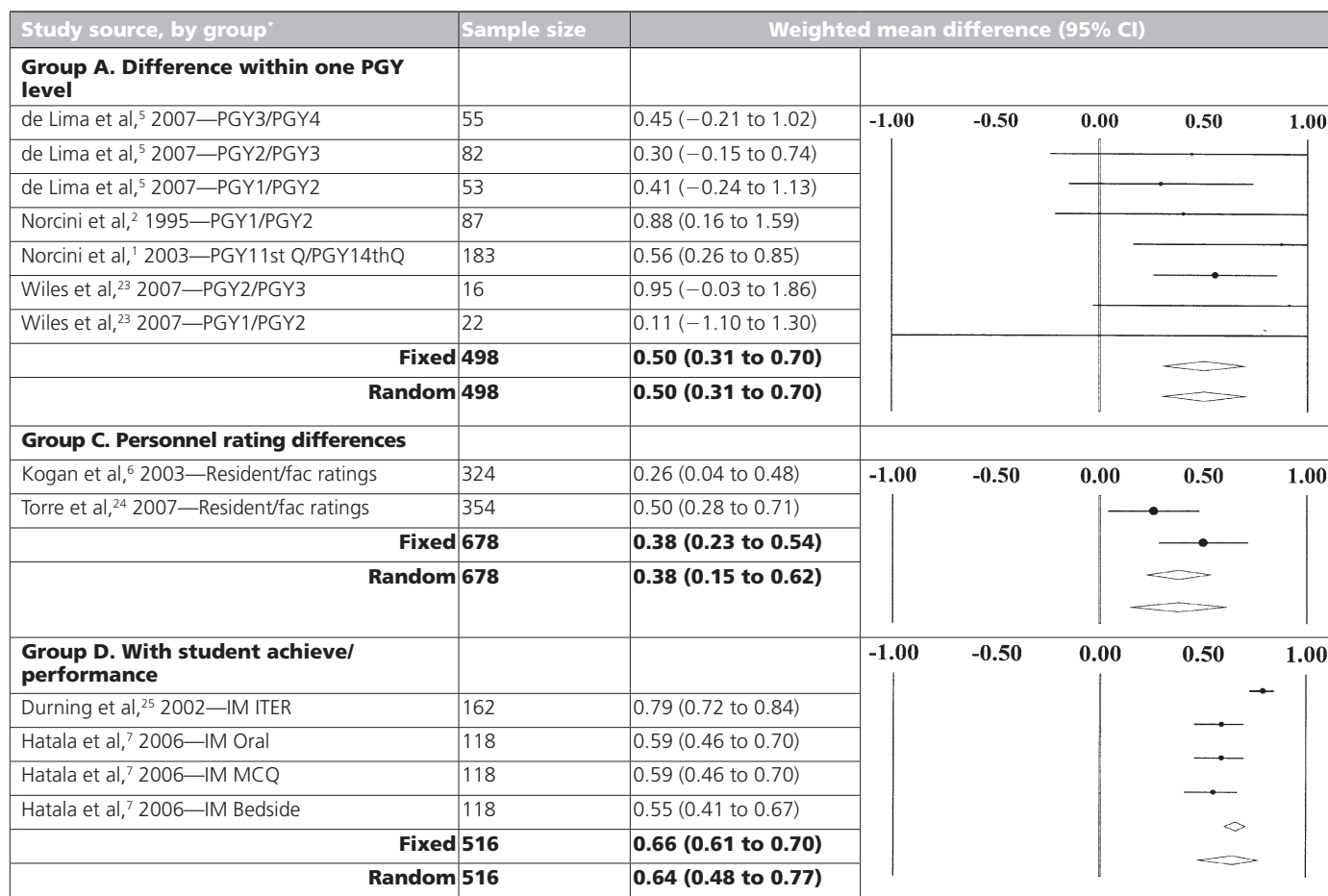
mini-CEX domain measure	Included studies (no. of outcomes)	Sample size	Group C. Personnel rating differences [†]	Included studies (no. of outcomes)	Sample size	Group D. With students' achievement/performance [‡]
1. Medical interview skills	2 (2)	678	0.38 (0.22–0.53)	3 (5)	505	0.51 (0.39–0.62)
2. Physical examination skills	2 (2)	678	0.35 (0.20–0.51)	3 (6)	536	0.58 (0.52–0.63)
3. Humanistic qualities/ professionalism	2 (2)	678	0.31 (0.15–0.46)	3 (5)	532	0.48 (0.32–0.62)
4. Clinical judgment	2 (2)	678	0.23 (0.04–0.50)	2 (4)	471	0.53 (0.32–0.69)
5. Counseling skills	2 (2)	678	0.50 (0.34–0.65)	1 (1)	10	0.85 (0.47–0.96)
6. Organization/efficiency	2 (2)	678	0.34 (0.19–0.50)	2 (4)	370	0.54 (0.46–0.62)
7. Overall clinical competence	2 (2)	678	0.38 (0.15–0.62)	2 (4)	516	0.64 (0.48–0.77)
Mean score on the mini-CEX	—	—	—	2 (9)	2,165	0.26 (0.16–0.35)

*Based on data from 11 mini-CEX studies from 1995 to 2012 identified in a meta-analysis of the research, to determine the construct and criterion validity of the mini-CEX as a measure of clinical performance.

[†]Effect size combined between personnel ratings (i.e., resident versus faculty, specialist versus consultant).^{6,24}

[‡]Effect sizes combined between standardized measures (i.e., American Board of Internal Medicine In-Training Evaluation Report, Royal College of Physicians and Surgeons of Canada Internal Medicine Exam, etc.) with students' achievement/performance exams.^{6,7,18,21,25}

Chart 1

Random- and Fixed-Effects Model Forest Plots for the Mini-Clinical Evaluation Exercise (Mini-CEX) "Overall Clinical Competence" Item*

*From the findings from 11 studies published between January 1995 and January 2012 identified in a meta-analysis of the published research to determine the construct and criterion validity of the mini-CEX as a measure of clinical performance. The effect size values were taken from the raw data reported for the outcomes in Table 1. PGY indicates postgraduate year. The Cochran *Q* test for heterogeneity shows significant overall heterogeneity between studies. A. $Q = 3.905, 6 df, P < .001$; C. $Q = 2406, 1 df, P < .01$; D. $Q = 18.882, 3 df, P < .001$ (Q statistic = a measure of heterogeneity as a function of the between studies, weighted squared deviations). For Group B, there were no data reported from the studies on item 7: overall clinical competence.

the findings outlined within the studies that were included in one or more of four group comparisons. As shown in Tables 2 and 3, the effect size calculations derived from the outcomes included in studies from Groups A, C, and D found a significant and “medium” combined effect size for each of the mini-CEX items or total mean score. As illustrated in the forest plot for the overall clinical competence item, not all reported outcomes on differences between a single year of residency training (Group A) were found to be statistically significant. When combined with the outcomes from four difference studies, however, we found that there is a significant total random-effects size of $d = 0.50$ (95% CI, 0.31–0.70). In support of the mini-CEX as a criterion-related measure in comparison with other

measures of clinical skills performance, we also found that there is a “large” combined random-effects size of $r = 0.64$ (95% CI, 0.48–0.77).

There are limitations to this meta-analysis. The quality of the meta-analysis study depends on the quality of the primary studies that we selected on the basis of the criteria we followed for inclusion. As we were interested in determining the construct- and criterion-related validity of the mini-CEX as a direct observation of clinical skill development, consistency in the use of the evaluation instrument from a research design perspective varied on the basis of who or what was being assessed (e.g., medical student, resident, performance-level ranking), the evaluators used (i.e., residents, faculty, specialists,

consultants), and whether or not the mini-CEX was being compared with other clinical skill measures (e.g., in-training evaluation reports, inpatient or outpatient write-ups, certifying exams). To overcome this concern, we used the more conservative random-effects size (weighted) analysis and test for heterogeneity between the studies by using the Cochran *Q* test. Although some of the studies had small sample sizes,^{17,21} such as 9 and 10, this was in part compensated by the 38 and 10 mini-CEX forms completed, respectively, by each of the participants in these studies. In an attempt to exercise additional control over the quality of the studies that were included, we selected articles that had been published in refereed journals. Further analyses based on potential moderator variables (e.g., sex,

age, residency discipline) were not possible because many of the included studies did not report such data.

The findings of this meta-analysis on the construct- and criterion-related validity of the mini-CEX show consistent and “medium” combined effect sizes for both mini-CEX items and total mean score. The introduction and use of the mini-CEX as a standard instrument for the assessment of medical students’ and residents’ clinical skill performance across a number of domains (e.g., medical interviewing, physical examination, counseling skills) are an important advance in the use and recognition of the direct observation method of clinical evaluation. Therefore, the mini-CEX has been adopted and used extensively as an instrument for the assessment of clinical skill performance in medical education programs in Canada, the United States, Europe, and other countries’ medical schools and residency programs.

The mini-CEX has the potential to be used frequently, even on a daily basis; however, the context of its use is restricted by the discipline (e.g., for general and specialty programs that have trainees encountering patients on a regular basis) and purpose of the evaluation (e.g., meets program expectations for the assessment of trainees’ clinical competencies). Consequently, there are some training programs where the use of the mini-CEX is simply not appropriate in many cases for some disciplines (e.g., anesthesiology, emergency medicine, pathology, radiology). Although the actual number of mini-CEX forms completed about a trainee will vary from program to program, it has been suggested that to obtain a generalizability coefficient of 0.88, only six encounters are required.²²

Medical educators are faced with the challenge of how to assess the in-training evaluation of medical students’ and residents’ clinical skills on the basis of patient encounters in a manner that is feasible, reliable, and valid. To date, the mini-CEX has proved to be a useful in-training assessment measure with clear evidence of construct- and criterion-related validity. Although the mini-CEX

is considered to be a useful assessment instrument, it should not be the only measure used to assess trainees’ clinical skill development. Other reliable and valid methods should be used in conjunction with the mini-CEX in the evaluation of clinical skills performance to overcome the limitation of using of a single measure.

Funding/Support: None.

Other disclosures: None.

Ethical approval: Not applicable.

References

- Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: A method for assessing clinical skills. *Ann Intern Med.* 2003;138:476–481.
- Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): A preliminary investigation. *Ann Intern Med.* 1995;123:795–799.
- Norcini JJ, Blank LL, Arnold GK, Kimball HR. Examiner differences in the mini-CEX. *Adv Health Sci Educ Theory Pract.* 1997;2:27–33.
- 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. *Adv Health Sci Educ.* 2011;16:189–199.
- Alves de Lima A, Barrero C, Baratta S, et al. Validity, reliability, feasibility and satisfaction of the mini-clinical evaluation exercise (mini-CEX) for cardiology residency training. *Med Teach.* 2007;29:785–790.
- 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–S35.
- Hatala R, Ainslie M, Kassen BO, Mackie I, Roberts JM. Assessing the mini-clinical evaluation exercise in comparison to a national specialty examination. *Med Educ.* 2006;40:950–956.
- Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA.* 2000;283:2008–2012.
- Cook DA, Beckman TJ, Mandrekar JN, Pankratz VS. Internal structure of mini-CEX scores for internal medicine residents: Factor analysis and generalizability. *Adv Health Sci Educ Theory Pract.* 2010;15:633–645.
- 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;85:1453–1461.

- Kogan JR, Hess BJ, Conforti LN, Holmboe ES. What drives faculty ratings of residents’ clinical skills? The impact of faculty’s own clinical skills. *Acad Med.* 2010;85(10 suppl):S25–S28.
- Margolis MJ, Clauser BE, Cuddy MM, et al. Use of the mini-clinical evaluation exercise to rate examinee performance on a multiple-station clinical skills examination: A validity study. *Acad Med.* 2006;81(10 suppl):S56–S60.
- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986;7:177–188.
- Cooper H, Hedges LV, Valentine JC, eds. *The Handbook of Research Synthesis and Meta-Analysis.* 2nd ed. New York, NY: Russell Sage Foundation; 2009.
- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis.* West Sussex, UK: John Wiley & Sons; 2009.
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Erlbaum; 1988.
- Holmboe ES, Huot S, Chung J, Norcini J, Hawkins RE. Construct validity of the miniclinical evaluation exercise (miniCEX). *Acad Med.* 2003;78:826–830.
- Ney EM, Shea JA, Kogan JR. Predictive validity of the mini-clinical evaluation exercise (mce): Do medical students’ mCEX ratings correlate with future clinical exam performance? *Acad Med.* 2009;84(10 suppl):S17–S20.
- Hatala R, Norman GR. In-training evaluation during an internal medicine clerkship. *Acad Med.* 1999;74(10 suppl):S118–S120.
- Hill F, Kendall K, Galbraith K, Crossley J. Implementing the undergraduate mini-CEX: A tailored approach at Southampton University. *Med Educ.* 2009;43:326–334.
- Boulet JR, McKinley DW, Norcini JJ, Whelan GP. Assessing the comparability of standardized patient and physician evaluations of clinical skills. *Adv Health Sci Educ Theory Pract.* 2002;7:85–97.
- 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. *Acad Med.* 2009;84(10 suppl):S113–S115.

References Cited in Tables and Charts Only

- Wiles CM, Dawson K, Hughes TA, et al. Clinical skills evaluation of trainees in a neurology department. *Clin Med.* 2007;7:365–369.
- Torre DM, Simpson DE, Elnicki DM, Sebastian JL, Holmboe ES. Feasibility, reliability and user satisfaction with a PDA-based mini-CEX to evaluate the clinical skills of third-year medical students. *Teach Learn Med.* 2007;19:271–277.
- 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;77:900–904.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1—Please confirm that “difference levels” (not “different levels”) is intended here.