

Feasibility, Reliability, and Validity of the Mini-Clinical Evaluation Exercise (mCEX) in a Medicine Core Clerkship

JENNIFER R. KOGAN, LISA M. BELLINI, and JUDY A. SHEA

ABSTRACT

Purpose. To determine the feasibility, reliability and validity of the mCEX when used to evaluate medical students' clinical skills in a medicine core clerkship.

Method. In 2002, students were required to complete nine mCEX during their medicine clerkship. Mean mCEX scores were correlated with exam scores and course grades.

Results. 89% of targeted mCEX were completed. The reproducibility coefficient for eight mCEX was .77. Mean mCEX scores were significantly correlated with exam scores ($r = .22$; $p = .004$), inpatient ($r = .43$; $p < .0001$), outpatient ($r = .35$; $p < .0001$), and final course grades ($r = .19$; $p = .014$).

Conclusions. These data support the feasibility, reproducibility, and validity of the mCEX in evaluating medicine clerkship students' clinical skills.

Graduating medical students must acquire the core clinical skills required for patient care and be able to demonstrate them on direct observation.¹ In this country, the two most common methods of student assessment during clinical rotations are written examinations and faculty and resident summative ratings.² Exams are an excellent measure of cognitive knowledge but scores do not necessarily correlate with clinical skills.³ Summative ratings fail to provide students with timely feedback regarding their clinical skills, and their validity is questionable when assessment is inferred rather than directly observed.³ Encounter cards and in-training evaluations have been used for formative assessment^{4,5} with varying degrees of success. Standardized patient examinations and OSCEs are reliable and valid clinical assessment tools but can be time-consuming and expensive. Therefore, identifying a feasible tool that promotes observation of medical students' clinical skills and can be used for summative assessment is important.

The mini-clinical evaluation exercise (mCEX) is a method of clinical skills assessment developed by the American Board of Internal Medicine (ABIM) to assess residents' clinical skills. Faculty observe and evaluate a resident during a focused new or follow-up patient encounter. The resident is evaluated along seven domains using a nine-point scale and then receives feedback. The mCEX is performed on multiple occasions with different patients and different observers. Studies have shown that the mCEX is brief, feasible and produces scores with adequate reproducibility when enough observations are made.⁶ Validity is suggested by correlations of the mCEX with other measures of residents' clinical competency.⁷

The mCEX also has been used as a feedback tool in the medicine clerkship.^{8,9} However, no study has assessed its test characteristics when used with medical students. Therefore, the purpose of this study was to determine the feasibility, reliability and concurrent validity of the mCEX when used as a tool to evaluate medical students' clinical skills in a medicine core clerkship. We hypothesized that when implemented with medical students, the mCEX would be feasible and reproducible, and concurrent validity would be supported by low but significant relationships with other assess-

ments and an observed increase in scores with increasing clinical experience.

Method

Participants

Core clerkships are organized into four, 12-week blocks. The Medicine Block includes a six-week inpatient and three-week outpatient internal medicine (IM) rotation, and a three-week family medicine (FM) rotation. This study includes all 165 students doing their IM clerkship between January and December 2002. The inpatient rotation uses four clinical sites. Faculty and residents are assigned in two- and four-week blocks, respectively. Therefore, students work with an average of three faculty and two residents during their inpatient rotation. The outpatient rotation uses 16 practices in which students work with one to three faculty.

mCEX Instrument and Implementation

mCEX booklets containing ten mCEX on duplicate pages were adapted with permission from the ABIM. Modifications included identifying if the evaluator was a resident or attending, eliminating the encounter complexity item and changing the focus item to "history," "physical exam" or "counseling." Students were evaluated on seven competencies (medical interviewing, physical examination, humanistic qualities/professionalism, clinical judgment, counseling, organization/efficiency and overall clinical competence) using a nine-point scale where 1 to 3 is unsatisfactory, 4 to 6 is satisfactory and 7 to 9 is superior. Evaluators could indicate if a particular skill was not observed. On each form, evaluators documented the number of minutes spent observing the student and providing feedback, and the evaluator and student rated their satisfaction with the mCEX using a nine-point scale (1 = low and 9 = high).

At each clerkship orientation, students received a mCEX booklet and were instructed to complete nine mCEX during the clerkship [three each from their inpatient attending(s), resident(s), and outpatient attending(s)]. Students were required to return their booklets the last day of the clerkship to pass the course, but the number of forms completed and actual mCEX evaluations did not contribute to final course grades. The study was approved by the Institutional Review Board.

Before implementation, faculty and residents were oriented to the rationale and use of the mCEX. Inpatient faculty were oriented during a mandatory quarterly meeting for faculty attending on inpatient services and residents during a mandatory annual "Residents as Teachers" in-service. Outpatient attendings were mailed an orientation packet, attended a faculty development retreat where the mCEX was discussed and were reminded to do the mCEX during annual site visits.

Validity Assessment

Each clerkship student takes the National Board of Medical Examiners (NBME) subject examination at the conclusion of the clerkship and receives a summative evaluation from every faculty member and resident with whom he/she worked for more than one week. The summative evaluation form contains 15 items in which stu-

TABLE 1. Mean (SD) Ratings for mCEX Competencies for Forms from Inpatient, Outpatient, and Resident Evaluators*

mCEX Competencies	Evaluators Mean (SD)			p Value for Trend	Effect Sizes†
	Inpatient Faculty (Max n = 340)	Outpatient Faculty (Max n = 363)	Resident (Max n = 434)		
Medical interviewing	7.4 (1.1)	7.5 (1.0)	7.7 (1.0)	.0001	.35
Physical examination	7.4 (1.0)	7.3 (1.1)	7.7 (1.1)	.0001	.43
Humanistic qualities/professionalism	8.0 (1.0)	8.1 (0.9)	8.2 (0.8)	.0001	.33
Clinical judgment	7.6 (1.0)	7.5 (1.1)	7.6 (1.0)	.49	.090
Counseling	7.7 (1.0)	7.6 (1.1)	8.0 (0.9)	.0001	.43
Organization/efficiency	7.5 (1.1)	7.6 (1.2)	7.8 (1.0)	.007	.26
Overall clinical competence	7.6 (1.0)	7.6 (1.0)	7.9 (0.9)	.0011	.26

*Cells in bold are significantly different, using Duncan's post hoc test.
 †Effect sizes are calculated for the highest to lowest mean.

dents are rated on a seven-point scale using behaviorally anchored items. A committee determines final inpatient grades by averaging the mean resident and mean faculty ratings in the context of written comments about the student. Outpatient grades are similarly determined by averaging faculty ratings and written comments. Students submit four patient write-ups that are graded by faculty and residents who are not those students' clinical preceptors. Final clerkship grades are a weighted mean of the student's inpatient (48%), outpatient (22%), exam (20%), and write-up (10%) grades.

Analyses

To determine feasibility, we examined the percentage of completed items and forms, average completion times and satisfaction ratings, and differences in ratings by different types of evaluators. Collapsing ratings within a student, interim correlations were computed and homogeneity was examined with Cronbach's alpha. A d-study followed a rater:student x item generalizability study. Validity was assessed with uncorrected Pearson correlations and ANOVAs with post-hoc comparison of means using Duncan's test looking at relationships between mCEX scores and block, exam scores, inpatient, outpatient, write-up, and final course grades.

Results

Analyses are based on the 162 students for whom we had mCEX evaluations and course grades. The total number of forms completed was 1,297 (89% completion rate): 340 forms were completed by inpatient faculty, 363 by outpatient faculty, and 434 by residents. This corresponded to a 70%, 75%, and 89% completion rate, respectively. One hundred sixty forms (12%) did not indicate the evaluator. The mean number of forms completed per student was 7.9 (range 2–10, median = 8). Competencies most frequently

evaluated included humanism (96% of forms), organization/efficiency (87% of forms), and physical examination skills (85% of forms). Medical interviewing, clinical judgment, and counseling were evaluated on 80%, 74%, and 46% of forms, respectively. Overall clinical competence was evaluated on 90% of forms. The mean observation and feedback time was 19 minutes (median = 15, range 1–180) and eight minutes (median = 5, range 0–122), respectively. The time spent observing students was significantly different for residents and faculty (p < .0001). The mean resident time of 22 minutes was significantly greater than the means for outpatient and inpatient faculty, 19 and 17 minutes, respectively. Mean times for feedback were not significantly different (p = .10). Evaluators' satisfaction with the mCEX was also significantly different (p < .0001). The mean satisfaction ratings were 7.7 (SD = 1.3) for inpatient faculty, 7.2 (SD = 1.8) for residents and 6.5 (SD = 2.1) for outpatient faculty. The mean student satisfaction rating was 7.0 (SD = 1.9). There was a significant difference in student satisfaction ratings depending on the evaluator (p = .03). Student satisfaction ratings were lower during encounters with outpatient faculty evaluators (mean = 6.8, SD = 1.8) than encounters with inpatient faculty (mean = 7.2, SD = 1.8) or residents (mean = 7.1, SD = 1.9). For the 1,297 forms, averaging ratings over all competencies within a form, mean ratings ranged from 4.5 to 9.0. On six items, there was a statistically significant difference in ratings assigned by residents and attendings (Table 1, p < .001), but effect sizes were generally small. In all cases, residents gave higher ratings than faculty. The lowest ratings were assigned to medical interviewing and physical examination and the highest to humanistic qualities.

For the 162 students, mean ratings averaged across forms ranged from 6.1 to 8.8. The overall mean was 7.7 (SD = 0.5). The homogeneity coefficient was .94. Correlations among individual

TABLE 2. Mean mCEX Scores for Students Obtaining Honors, High Pass, and Pass Grades on Different Summative Assessments*

	Grade				
	Self Exam	Write-up	Inpatient	Outpatient	Final Course
Honors	7.9 (0.6) n = 53	7.8 (0.6) n = 81	8.1 (0.4) n = 58	7.9 (0.6) n = 69	8.0 (0.4) n = 64
High pass	7.7 (0.6) n = 51	7.7 (0.6) n = 65	7.5 (0.6) n = 95	7.6 (0.6) n = 74	7.5 (0.6) n = 85
Pass	7.6 (0.7) n = 55 p = .0434	7.4 (0.8) n = 16 p = .0448	7.2 (0.9) n = 8 p = .0001	7.3 (0.6) n = 15 p = .0001	7.5 (0.8) n = 9 p = .0001

*Students who received grades of fail or incomplete were eliminated from analyses secondary to small numbers. **Bold** indicates differences that are significantly different with post hoc Duncan's test.

competencies ranged from .53 to .78 and correlations between the six competencies and overall clinical competence ranged from .69 to .87. Reproducibility estimates for varying numbers of raters per student were .62 for four raters, .71 for six raters, and .77 for eight raters.

There were significant correlations (uncorrected) between mean mCEX scores and exam scores ($r = .22$; $p = .004$), write-ups ($r = .17$, $p = .035$), inpatient ($r = .43$; $p < .0001$), outpatient ($r = .35$; $p < .0001$), and final course grades ($r = .19$; $p = .014$). Students who received "Honors" on their exam, inpatient, outpatient, and final course grade had higher mean mCEX scores than those students who received a "Pass" (Table 2). As expected, mean mCEX ratings increased over the course of the year ($p < .0001$). Scores were significantly lower for the first block (mean = 7.4, SD = 0.6), but similar though steadily increasing through the second (mean = 7.7, SD = 0.6), third (mean = 7.8, SD = 0.6) and fourth blocks (mean = 7.9, SD = 0.6).

Discussion

Acquiring core clinical skills requires feedback that can only be provided in the context of direct observation. Such observation is also necessary to document clinical competence. Frequent, direct observation of student-patient interactions can be challenging on an acutely ill, inpatient service or in a busy, outpatient practice. Commonly, assessment of clinical skills is inferred through evaluators' recollections of students' case presentations which may not accurately reflect students' clinical skills.²

This study demonstrates that the mCEX is feasible to use in an inpatient and outpatient medicine clerkship for formative assessment. Students were assessed multiple times in brief encounters that were favorably perceived by students and evaluators. Additionally, the mCEX has reasonable reproducibility with eight mCEX having a reproducibility coefficient of .77. The significant correlations between mCEX scores and clerkship summative ratings, exam scores and write-up grades suggest concurrent validity. The low magnitude of the observed correlations, albeit certainly lower than the true correlations, suggest the mCEX may be measuring distinct skills. It is also reassuring that "Honors" students had significantly higher scores than students assigned a "Pass" and mCEX performance improved over the course of the year. Similar to prior studies of in-training evaluation, residents were more lenient evaluators than faculty,⁴ an important consideration if the mCEX is used for summative assessment. The higher ratings by residents suggest the source of ratings needs to be balanced across students. Making the process more acceptable to outpatient attendings also would be beneficial.

There are several limitations. First, this is a single institution study. Generalizability to students at other institutions will be

necessary. Secondly, the correlation between mCEX scores and students' summative evaluations is not unexpected since many faculty and residents who completed the mCEX also submitted summative evaluations, usually a few weeks after the mCEX encounter. While mCEX scores did not contribute to students' final course grades, it is likely that the observers' summative evaluations were influenced by mCEX observations. Future analyses will need to eliminate common raters. Regardless, validity is still suggested by the correlation of mCEX scores with exam and write-up grades where no overlap in evaluators existed. Thirdly, internal consistency was extremely high suggesting considerable redundancy in the assessments, along the lines of a generic global assessment. Additionally, evaluators were not formally "trained" to use the form. Future studies might investigate strategies to encourage better use of the scale, perhaps by adding behavioral anchors that reflect student performance objectives. Finally, since this study did not have a control group, the effect of the mCEX on learning, performance and feedback cannot be determined.

We have demonstrated that the mCEX can be used in a medicine clerkship for both formative and summative assessment of medical students' clinical skills. Whether it will perform similarly well in other core clerkships such as pediatrics, obstetrics/gynecology or surgery needs to be determined.

Correspondence: Jennifer R. Kogan, MD, 3701 Market Street-640, Philadelphia, PA 19104; e-mail: (jennifer.kogan@uphs.upenn.edu).

References

1. The Medical Schools Objectives Writing Group. Learning objectives for medical student education. Report I of the medical schools objectives project. *Acad Med.* 1999;74:13-8.
2. Kassebaum DG, Eaglen RH. Shortcomings in the evaluation of students' clinical skills and behaviors in medical school. *Acad Med.* 1999;74:842-9.
3. Stillman PL, Regan MB, Swanson DB, et al. An assessment of the clinical skills of fourth year students at four New England medical schools. *Acad Med.* 1990;65:320-6.
4. Hatala R, Norman GR. In-training evaluation during an internal medicine clerkship. *Acad Med.* 1999;74:S118-20.
5. Turnbull J, MacFadyen J, van Barneveld C, et al. Clinical work sampling: a new approach to the problem of in-training evaluation. *JGIM.* 2000;15:556-61.
6. Norcini JJ, Blank LL, Arnold GK, et al. The mini-CEX (clinical evaluation exercise): A preliminary investigation. *Ann Intern Med.* 1995;123:795-99.
7. Durning SJ, Cation LJ, Markert RJ, et al. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Acad Med.* 2002;77:900-4.
8. Hauer KE. Enhancing feedback to medical students using a mini-CEX (clinical evaluation exercise). *Acad Med.* 2000;75:254.
9. Kogan JR, Bellini LM, Shea JA. Implementation of the mini-CEX to evaluate medical students' clinical skills. *Acad Med.* 2002;77:1156.