

Do faculty show the 'halo effect' in rating students compared with standardized patients during a clinical examination?

D Lie, J Encinas, F Stephens, M Prislin

Citation

D Lie, J Encinas, F Stephens, M Prislin. *Do faculty show the 'halo effect' in rating students compared with standardized patients during a clinical examination?* . The Internet Journal of Family Practice. 2009 Volume 8 Number 2.

Abstract

BackgroundWe compared standardized patient (SP) and faculty ratings of student performance to test the hypothesis that faculty systematically overestimate student clinical skills compared with SPs. **Methods**We conducted a study of three successive classes of second year medical students (n=303). Students were assessed by SPs during an Objective Structured Clinical Examination and faculty simultaneously rated student performance using the modified mini-clinical evaluation exercise (mini-CEX). The Patient Physician Interaction scale (PPI), a communication scale designed for use by SPs and the mini-CEX measure designed for use by faculty were each analyzed for internal consistency reliability. Within the SP and faculty measures scores for similar skills of history-taking, physical examination, humanism/communication and global score were analyzed in aggregate for concordance. Range and patterns of SP and faculty scores were examined and compared. **Results**The PPI and mini-CEX each demonstrated high internal consistency (Cronbach's α .86, .89). Correlation between clinical skill (history, examination, humanism/communication and global) score as rated by SPs and faculty was modest (Pearson's r 0.213 to 0.288). SPs and faculty both used the entire range of scores within each measure with a normal distribution. **Conclusions**The PPI and mini-CEX have high internal consistency reliability. There is only modest correlation between SP and faculty ratings of student clinical skills. Differences should be recognized when giving student immediate feedback on their performance in a standardized examination. Faculty rate students without a 'halo' effect.

INTRODUCTION

Faculty feedback to medical students regarding their performance in clinical encounters contributes to learning and improvement of clinical skills¹⁻². Observations of student performance are an integral part of student evaluation in clerkships and can be achieved in different ways, such as observed history and physicals in the inpatient or outpatient³⁻⁴ setting, or in a standardized clinical examination⁵⁻⁷. However, constraints of limited faculty time and busy practices often limit opportunities for observation with feedback in actual or standardized encounters. Among a myriad of assessment tools available⁸ the mini-clinical evaluation exercise or mini-CEX was developed and validated⁹⁻¹⁰ with medical students and residents³ to address some of the faculty time constraints. The mini-CEX allows observation, evaluation and feedback by faculty that focuses separately on the skills of history taking, physical examination, clinical reasoning, humanistic qualities and overall clinical competence. The mini-CEX has been

adapted for use in medical student¹¹⁻¹³ and residency¹⁴ settings for summative as well as formative evaluation by faculty. Many preclinical Doctoring courses and clerkships now incorporate the Objective Structured Clinical Examination (OSCE) or other standardized patient scenarios^{5,7,15} to assure that students are not only observed and rated¹⁶ but given feedback by standardized patients (SPs) and faculty¹⁷. In such settings, faculty or course directors also use SP scores to provide feedback to learners.

A comprehensive review of assessment tools used for feedback in clinical teaching¹⁸ suggested that characteristics of teacher evaluations vary across educational settings and that future studies should focus on narrowly defined study populations. It is unclear how well measures of performance by trained SPs correlate with faculty ratings of similar skills for students, since the two types of observations are often independent of one another and occur in separate settings (standardized vs. actual) using different scales with different construct validity. It has been suggested that faculty may

exhibit the 'halo effect' and tend to overestimate performance when asked to assess the skills of trainees well known to them, applying their global impressions rather than objective ratings of specific behaviors¹⁹⁻²¹, while SPs trained to use checklists observe specific verbal and nonverbal behaviors in standardized encounters are more likely to be "objective" in their ratings, relying less on global impressions²². Providing consistent feedback based on faculty and SP ratings becomes a challenge when the ratings or scores from SPs and faculty diverge, especially when faculty ratings are consistently higher than SP ratings. Despite this potential for discordance, there is a paucity of studies examining the correlation between ratings and rating patterns produced by faculty and SPs in the same clinical encounters^{18, 20}.

We therefore conducted a study to test the hypothesis that faculty overestimate their students' clinical performance resulting in discordance with SP ratings. Our purpose was first to establish the validity and reliability of two commonly used validated assessment tools: the faculty-completed mini-CEX¹⁰ and the patient-completed Patient Physician Interaction Scale or PPI¹⁵ in a four station OSCE for second year medical students. Our second objective was to examine the distribution pattern of ratings by SPs and by faculty. In addition we examined the correlation between faculty and SP raters for four individual skill domains within the modified mini-CEX. Rather than asking faculty to use the exact same scales for clinical performance as trained SPs we used different rating measures for SP and faculty to reflect real world clinical teaching where faculty do not generally use case-specific SP checklists to rate student skills. The mini-CEX is not case-specific and was constructed and validated for use by faculty for use in most clinical encounters, and not for use by SPs. In addition, SP measures are considerably longer to complete, require time-consuming training, and impractical for faculty to complete without case-specific training. The institutional review board of the University approved the study.

METHODS

SETTING AND CURRICULUM

The study was conducted at one US medical school with class sizes of 92 to 104. Three consecutive classes of students were tested for their clinical skills using the same four-station OSCE at the end of 18 months of a required longitudinal Doctoring course. The primary goal of the 180-hour Doctoring course was to introduce basic interviewing,

physical examination, communication, and clinical reasoning skills to prepare students for third year clerkships. The course started in the fall of the first year and continued to the spring of the second year, and took place concurrently with other required basic science courses including Anatomy, Physiology, Pharmacology, and Pathology. The course comprised eight consecutive organ-system-based clinical cases (matched to teaching in the basic science courses) with opportunities for structured observed interview and physical examination of standardized and real patients in weekly two-hour problem-based learning (PBL) sessions. Students were taught in groups of four to six by a pair of faculty, one a physician, and the other a related health professional (e.g. a psychologist, nurse, or social worker). Small group teaching was standardized using a structured schedule with learning objectives, SP cases, and homework assignments, supported by intensive faculty development, which occurred monthly. PBL sessions were videotaped and feedback on interviewing, physical examination skills, communication, and group participation was given to students. Small group teaching was supplemented by teaching of content themes relevant to the cases, weekly half day preceptorships with practicing community physicians during the second year, and required readings from textbooks. Student evaluations consisted of: individual assessments by PBL instructors twice a year, , assessments by community preceptors, a written test, and the final OSCE given in the second year. Physician PBL instructors had at least 32 hours of face-to-face teaching contact with their small groups and evaluated students weekly on their homework and class performance as well as meeting individually with each of their students once every 2 to 3 months.

The OSCE was administered during the Winter Quarter of the second year to the entire class over a three-week period on eight half days at the medical school's Clinical Skills Training Center with a level of test security equivalent to other high stakes clinical examinations offered at the school.

STUDY PARTICIPANTS

Student participants were second year medical students. Faculty raters were instructors in the PBL groups for the students. SPs were trained actors with prior experience as raters in high stakes clinical practice examinations.

OSCE CASES

The goal of the OSCE was to determine that students had

basic interviewing, physical examination and clinical reasoning skills appropriate to the course goals and content. Four cases were developed and pilot-tested in the year prior to the study to address learning in the skill domains and organ systems taught. Each station was 25 minutes long, with a 10-minute inter-station exercise for writing up the case or to answer case-specific questions. The cases were: a 45-year old businessman presenting with atypical chest pain with a differential diagnosis of ischemic heart disease/angina, gastroesophageal reflux and pleurisy; a 60-year old woman with transient neurological symptoms and vascular disease; a 30-year old man or woman with acute back pain secondary to muscle strain from work; and a 23-year old male college student with acute right lower quadrant abdominal pain with a differential diagnosis of appendicitis, urinary infection, kidney stone or testicular torsion.

During the OSCE, faculty observed the encounters live from a television monitor outside the room. They were provided with written instructions for the modified mini-CEX to rate each skill domain based on their impression of skill appropriate for a second year medical student and to provide written narratives to support their ratings. All faculty were introduced to the modified mini-CEX during the Doctoring course, and were trained in the use of anchors for the rating scales within the mini-CEX. We chose to use the mini-CEX for the OSCE because of faculty familiarity with the tool and for consistency with clerkships which also use the mini-CEX for student assessment.

RATING OF STUDENT CLINICAL SKILLS

SP ratings: History taking and physical examination skills were assessed using case-specific checklists.

Communications skills were measured using the same seven-item, Likert type five-point, rating scale (1=marginal/low to 5=outstanding), and the PPI¹⁵ for all four cases (see table 1 for individual PPI items). The maximum score for the PPI for each case was 35 points. Overall patient satisfaction (OS) was measured using a single-item, 5 point scale with a range from 1= highly dissatisfied to 5=highly satisfied.

Faculty ratings: All faculty raters completed the mini-CEX^{3,4,23} once on each student observed at one station. The skills rated were history taking, physical examination, communication, humanism and overall clinical competence (see table 1 for rating domains), with a rating scale from 1=unsatisfactory to 5=exceptional.

Internal consistency reliability: The PPI was the common communication skills measure used by different SPs across the four OSCE cases and its internal consistency reliability was assessed. The entire modified mini-CEX (5 skill domains) was assessed for its internal consistency reliability (table 1) among the different faculty raters.

Comparison of individual SP and faculty skill measures: Four different clinical skills were compared for SP and faculty ratings. The SP history-taking score or Hx was the sum of all history-taking items (binary scoring yes/no or 1/0 for each item) for each case corrected to a percentage score. The matching faculty history-taking score was the single score for question one (CEX1) of the mini-CEX. The SP physical examination (PE) score or Px was a sum of the individual behaviors (binary scoring yes/no or 1/0 for each item) for that station corrected to 90% plus the Fundamentals of Physical Examination Behaviors (FPEB) score corrected to 10% for a total of 100%. The FPEB consisted of 2 questions (binary yes/no or 1/0 for each item) asking if students explained the examination and if they respected the modesty of the patient. The matching faculty PE score was the score for question two (CEX2) on the mini-CEX. The SP 'humanistic quality/communication score' was the sum of the PPI score, with points totaled (maximum score for summed PPI =35). The equivalent faculty rating was the score for question four (CEX4) on the mini-CEX. The SP overall satisfaction (OS) item for each station ('please rate the overall performance of the student') was matched to the score for question five on the modified mini-CEX asking about the overall performance of the student (CEX5). Question three on the modified mini-CEX (CEX3), asking about the clinical reasoning skill of students, was not matched to any SP rating, as SPs were not asked to rate students' clinical reasoning skills.

STATISTICAL ANALYSIS

Statistical analysis was performed using the software package JMP version 7.0.1 (SAS Institute Inc., Cary NC). Because analysis by student class showed no differences in means and SD among the 3 classes, the data was analyzed in aggregate. The means and standard deviations for the modified mini-CEX and PPI were calculated. Multivariate analysis was used to calculate Cronbach's α to establish internal consistency reliability for individual items and combined items of the modified mini-CEX (faculty raters) and the PPI (SP raters).

Pearson r correlations between each mini-CEX item (1, 2, 4,

and 5) and the corresponding SP measure (described above) were calculated. CEX 1, 2, 4 and 5 (using the 5-point Likert scale) were compared with each relevant variable: Hx% with CEX1, Px 90%+ FPEB10% with CEX2, PPI total score with CEX4, and OS with CEX5.

Subgroup analysis was performed to examine the correlations using Spearman ρ , by student gender, student class, faculty gender, and clinical case.

RESULTS

STUDY PARTICIPANTS

The OSCE was conducted in the winter of 2005, 2006 and 2007. Each of the three classes (N=303 total) was similar in demographics. Mean age was 23 years, 50% were male, and distribution of ethnicity was as follows: 40-50% white, 40% Asian, 10% Hispanic, 1% black, and 10% did not self-categorize. The majority of students were science majors and residents of California.

There were 34 physician faculty raters (16 male) from primary care settings of Family Medicine (n=16), Internal Medicine (n=15) Pediatrics (n= 2), and Psychiatry (n=1). Faculty ranks ranged from Assistant to Full Professor. Individual faculty experience teaching in the Doctoring course ranged from 3 to 10 years. Each faculty rated 3 to 17 students. The majority of faculty (25 of 34) taught students during all 3 years of the study.

Twenty-four SPs performed in the three OSCEs over 3 years and each performed in the same case. There were two SPs for each clinical case who were trained concurrently for 4 hours each, using videotapes of student performance as "gold standards" for rating. Returning SPs were retrained each year for 2 hours with review of the case and rating standards.

OVERALL STUDENT PERFORMANCE

Mean (SD) scores out of 5 for modified mini-CEX clinical skill domains assessed by faculty ranged from the lowest mean score of 3.04 (\pm 0.76) for physical examination, to the highest mean score of 3.52 (\pm 0.58), for humanistic qualities (Table 1), equivalent to percentage scores of 60.80% to 70.40% respectively. Mean (SD) SP percentage scores ranged from a low of 52.90% (\pm 16.99%) for physical examination, to a high of 74.85% (\pm 17.72%) for history taking. The mean score for the PPI summed total was 25.45 (\pm 3.02), out of a maximum of 35 points and the mean OS score from SPs was 3.69 (\pm 0.58), out of a maximum of 5

points (figure 1).

Mean scores for each case and by skill domain assessed by SPs did not improve across the three years of the study (data not shown), suggesting good test security across the years of OSCE administration.

INTERNAL CONSISTENCY RELIABILITY OF THE PPI AND THE MODIFIED MINI-CEX

The overall Cronbach's α for the PPI was 0.86. The overall Cronbach's α for the entire mini-CEX was .89 (table 1).

FACULTY-SP RATINGS – SCORE DISTRIBUTION AND CONCORDANCE

Both faculty and SPs used the full range of the individual scales (see figure 1 for distribution of scores for measures) with scores demonstrating a normal distribution. The Pearson correlation coefficient r between the modified mini-CEX and the equivalent SP measures were 0.214 for CEX1 and Hx percentage score, 0.288 for CEX2 and Px (90% physical examination + 10%FPEB) score, 0.213 for CEX4 and the PPI summed total score, and 0.245 for CEX5 and overall satisfaction SP score (Figure 1).

Subgroup analysis (table 2) showed that the Spearman ρ was lower in all four modified mini-CEX domains for female compared with male students. The range was 0.10 to 0.18 for females and 0.23 to 0.33 for males. There were otherwise no consistent patterns of correlation by case, student class, or faculty gender (Table 2). However, one of the four cases (abdominal pain, case 3) showed a higher range of variability in concordance (Spearman ρ -0.032 to +0.473) than the other 3 cases (table 2) which may be due to the larger number of SPs used for this compared with the other three cases. When data was examined by individual faculty, no single faculty demonstrated a tendency to score in a narrow low or high range (data not shown).

Do faculty show the 'halo effect' in rating students compared with standardized patients during a clinical examination?

Figure 1

Table 1: Modified Mini-CEX and PPI Scores and Variances - University of California, Irvine, School of Medicine 2005-2007

Scale	Item Questions	Mean (SD)	Item Correlation Coefficient (ICC)
MODIFIED MINI-CEX	Combined total Mini-CEX (Cronbach's $\alpha = 0.89$)	15.74 (3.04)	-
	1. Clinical Skills - History	3.25 (0.68)	0.86
	2. Clinical Skills - Physical Examination	3.04 (0.76)	0.87
	3. Clinical Judgment and Synthesis (Based on assessment & plan and oral presentation)	3.15 (0.69)	0.85
	4. Humanistic Qualities	3.52 (0.58)	0.90
	5. Overall Clinical Competence as a Student Physician (as demonstrated in this focused exercise)	3.27 (0.66)	0.84
PPI	Combined total PPI scale (Cronbach's $\alpha = 0.86$)	25.45 (3.01)	-
	1. Appeared professionally competent	3.71 (0.59)	0.84
	2. Effectively gathered information	3.64 (0.63)	0.84
	3. Listened actively	3.70 (0.51)	0.84
	4. Established personal rapport	3.70 (0.60)	0.84
	5. Appropriately explored my perspective	3.57 (0.53)	0.84
	6. Addressed my feelings	3.63 (0.56)	0.84
	7. Met my needs	3.51 (0.68)	0.85

LEGEND

- Each item on the modified mini-CEX used a 5-point rating scale, with a value of 1 = "unsatisfactory" performance, and 5 = "exceptional" performance, respectively.
- Each item on the PPI used a 5-point rating scale, with a value of 1 = "marginal/low" performance, and 5 = "outstanding" performance, respectively.
- Faculty completed the modified mini-CEX and the standardized patient completed the PPI.

Abbreviations:

CEX (Clinical Evaluation Exercise)
PPI (Patient Physician Interaction)

Figure 2

Table 2: Correlations (Spearman ρ) Between Faculty and SP Measures by Subgroup Analysis -University of California, Irvine School of Medicine 2005-2007

Characteristics	Total, n(%)	Spearman ρ (rho)			
		CEX1 vs 100% Hx	CEX2 vs 90% Px + 10% FPB	CEX4 vs PPI Total	CEX5 vs OS
Student Gender (N=303)	-	-	-	-	-
Male	156 (51.5)	0.232	0.326	0.195	0.283
Female	147 (48.5)	0.096	0.184	0.171	0.095
Student Class (N=303)	-	-	-	-	-
1 ^a	97 (32.0)	0.211	0.298	0.251	0.323
2 ^b	100 (33.0)	0.004	0.235	0.313	0.263
3 ^c	106 (35.0)	0.228	0.230	-0.001	0.108
Faculty Gender (N=34)	-	-	-	-	-
Male	16 (45.0%)	0.221	0.219	0.142	0.170
Female	18 (55.0%)	0.119	0.292	0.233	0.231
SP Cases (N=303)	-	-	-	-	-
1 ^d	78 (25.7)	0.093	0.258	0.121	0.268
2 ^e	75 (24.8)	0.164	0.226	0.179	0.003
3 ^f	78 (25.7)	-0.032	0.473	0.160	0.305
4 ^g	72 (23.8)	0.360	0.024	0.272	0.218

LEGEND

- a 2005 (Year of exam)
- b 2006 (Year of exam)
- c 2007 (Year of exam)
- d Neurological (Case)
- e Chest Pain (Case)
- f Abdominal Pain (Case)
- g Back Pain (Case)

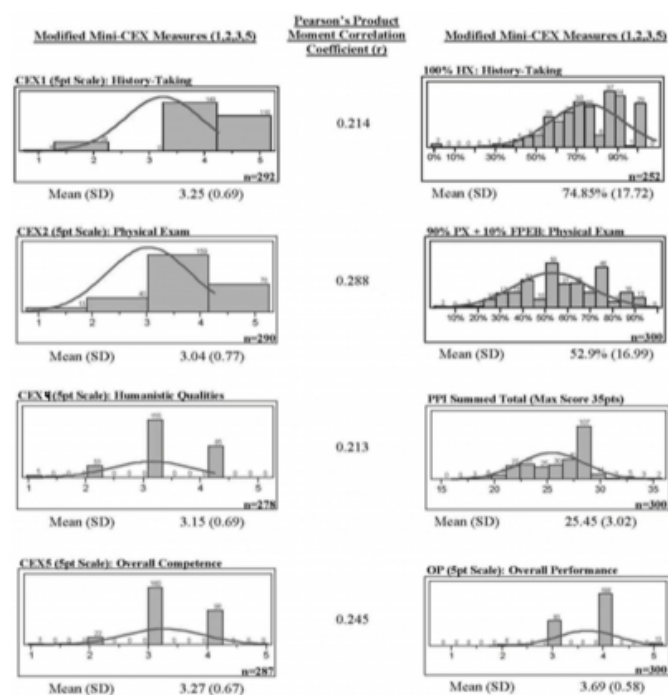
• Highest and lowest Spearman ρ of each paired comparison denoted in bold

Abbreviations:

CEX (Clinical Evaluation Exercise)
Hx (History)
Px (Physical Exam)
FPB (Fundamentals of Physical Exam Behaviors)
PPI (Patient-Physician Interaction)
OS (Overall Satisfaction)
SP (Standardized Patient)

Figure 3

Figure 1: Correlations (Pearson r) and Variances of the Modified Mini-CEX and Matched SP Measures University of California, Irvine School of Medicine



LEGEND

Abbreviations:
CEX (Clinical Evaluation Exercise)
Hx (History)
Px (Physical Exam)
FPB (Fundamentals of Physical Exam Behaviors)
PPI (Patient Physician Interaction Scale)
SP (Standardized Patient)

Note - Number for each measure is lower than 303 due to some missing data

DISCUSSION AND CONCLUSION

We examined the internal consistency reliability of two measures (one for use by SPs and one for faculty) used for clinical skills assessment, in second year medical students performing in an OSCE for 3 consecutive classes of students. We confirmed prior reports of reliability and validity of both measures, with Cronbach's α of .86 for the PPI and 0.89 for the mini-CEX. We also tested the concordance of faculty and SP measures of equivalent clinical skills, for history taking, physical examination, communication/humanistic qualities and overall skills across four stations, using faculty raters who taught the students and had experience and training using the modified mini-CEX. As expected²⁴, student performance by faculty and SP rating was lowest in the area of physical examination, consistent with our previous observations of second year student performance. Faculty and SPs diverged in the area of highest mean scores: faculty scores were highest for humanistic qualities (CEX4) while SP scores were highest in

the area of history taking (Hx).

Although the concordance between SP and faculty raters was modest with a range of 0.214 to 0.288 for the 4 different skills examined, both sets of raters in aggregate showed a consistent normal distribution pattern of rating across the entire scales of their respective assessment tools. There was no suggestion of a consistent “halo effect” in either group of raters. The concordance between faculty and SPs was somewhat lower for female compared with male students but no other pattern related to class, clinical case or faculty gender was observed.

Strengths of our study are the large number of students assessed with almost 100% of data available for analysis. The same curriculum was delivered for each of the 3 groups of students. Student demographics were similar across the 3 years allowing aggregate analysis. The total number of faculty and SPs was small and consistent across the study years. The majority of faculty was experienced in PBL instruction and clinical skills rating and all had similar contact hours with their students. The cases and SPs used were the same across the three years of the study. The number of students was sufficiently high to permit meaningful subgroup analysis.

There are some limitations to this study. First, we used aggregate data with means representing SP and faculty scores instead of individual paired comparisons. This was because each student-SP encounter was observed and rated by a single unique faculty and a unique SP. However, we had sufficient observations of each faculty and still found no “outlier” faculty with tendencies to score narrowly at the upper or lower end of the modified mini-CEX scale²⁵. Second, we converted SP scores to match skill domains for the modified mini-CEX but the skills measured using this conversion are not equivalent in construct. For example, the sum of the PPI items do not necessarily equate to “humanistic qualities” as measured by the modified mini-CEX4. The reason we chose this strategy, instead of having faculty use the same rating scales as the SPs, was to reflect the “real world” conditions of faculty assessment of clinical skills. In clinical precepting by faculty, case-specific checklists for encounters are not generally adopted because of the unpredictability of each encounter and each

case type. This mismatch in construct likely accounts for the low to modest concordance we found. Lastly, this study was conducted at one medical school and results may not be generalizable across other schools with different faculty and SP characteristics and experience.

The finding that this diverse (by gender, specialty and teaching experience) group of faculty showed no overall tendency to overestimate skills of students known to them is reassuring and may be explained by faculty development for this particular faculty group. Our study suggests that agreement between faculty and SPs about student skills in an observed encounter is not high when different rating scales are used, which supports prior literature differentiating global ratings and checklist ratings²⁶⁻²⁷. It is likely that faculty in our study adopted a global approach for each skill, while SPs relied on itemized behavior-specific checklists for all the skills except for the overall score. Correlation between faculty and SP was not high ($r=0.245$) even for overall performance comparing the single SP global item with the modified mini-CEX5. In clinical teaching, immediate feedback after a clinical encounter is often much appreciated by students, and carries higher impact compared with delayed feedback²⁸⁻²⁹. However, when more than one source of assessment is available for the same encounter, conflicting reports have the potential to be detrimental to learning³⁰⁻³¹. Based on our findings, we recommend that, when giving real-time feedback to students based on performance in a standardized examination using SPs, attention should be paid to SP ratings before faculty feedback is given, to optimize the consistency and impact of the feedback. Differences in scoring between SP and faculty should be acknowledged and the potential reasons for discordance explored with students. Future studies will focus on identifying reasons for, and developing appropriate faculty development strategies to address potential discrepancies in patient vs. faculty ratings to optimize student learning.

Word count for main text: 3318

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the dedication and commitment of the many volunteer faculty who taught in the Doctoring course, our medical students, Sue Ahearn, Charlotte Fesko and the Clinical Skills Center for maintaining the highest standards of case development. This project is partially supported by funding from an award to

DAL from the National Institutes of Health (NIH), National Heart, Lung and Blood Institute, K07 HL079256-01 (2004-10) and from the Bureau of Health Professions, grants D55HP03362-01-00 (2004-9). The manuscript contents are solely those of the authors and do not reflect on views of the National Institutes of Health.

References

- Ende J. Feedback in clinical medical education. *JAMA*. Aug 1983;250(6):777-781.
- Salerno SM, Jackson JL, O'Malley PG. Interactive faculty development seminars improve the quality of written feedback in ambulatory teaching. *J Gen Intern Med*. Oct 2003;18(10):831-834.
- Durning SJ, Cation LJ, Market RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Acad Med*. 2002;77(9):900-904.
- Norcini JJ. The mini clinical evaluation exercise (mini-CEX). *The Clinical Teacher*. Jun 2005;2(1):25-30.
- Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ*. Jan 1979;13(1):41-54.
- Van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: The state of the art. *Teach Learn Med*. 1990;2:58-76.
- Makoul G, Altman M. Early assessment of medical students' clinical skills. *Acad Med*. Nov 2002;77(11):1156.
- Zubair Amin Z, Seng CY, Eng KH. *Practical Guide to Medical Student Assessment*. Pte. Ltd.; 2006:106.
- Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): A preliminary investigation. *Ann Intern Med*. Nov 1995;123(10):795-799.
- Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: A method for assessing clinical skills. *Annals of Internal Medicine*. Mar 2003;138(6):476-481.
- Hauer KE. Enhancing feedback to students using the mini-CEX (clinical evaluation exercise). *Acad Med*. May 2000;75(5):524.
- Kogan JR, Bellini LM, Shea JA. Implementation of the mini-CEX to evaluate medical students' clinical skills. *Acad Med*. Nov 2002;77(11):1156-1157.
- Kogan JR, Bellini LM, Shea JA. Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mini-CEX) in a medicine core clerkship. *Acad Med*. Oct 2003;78(10 Suppl):S33-5.
- Kogan JR, Hauer KE. Brief report: Use of the mini-clinical evaluation exercise in internal medicine core clerkships. *J Gen Intern Med*. May 2006;21(5):501-502.
- Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. *Acad Med*. Apr 2001;76(4):390-393.
- Newble D. Techniques for measuring clinical competence: Objective structured clinical examinations. *Med Educ*. Feb 2004;38(2):199-203.
- Petrusa ER. Status of standardized patient assessment: Taking standardized patient-based examinations to the next level. *Teach Learn Med*. Jan 2004;16(1):98-110.
- Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? A review of the published instruments. *J Gen Intern Med*. Sep 2004;19(9):971-977.
- Norcini JJ, Blank LL, Arnold GK, Kimball HR. Examiner differences in the mini-CEX. *Adv Health Sci Educ*. 1997;2(1):27-33.
- 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*. Oct 2006;81(10 Suppl):S56-60.
- Barman A. Critiques on the Objective Structured Clinical Examination. *Ann Acad Med Singapore*. Sep 2005;34(8):478-482.
- Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med*. Sep 1998;73(9):993-997.
- Holmboe ES, Huot S, Chung J, Norcini J, Hawkins RE. Construct validity of the miniclinical evaluation exercise (mini-CEX). *Acad Med*. Aug 2003;78(8):826-30.
- Prislin MD, Giglio M, Lewis EM, Ahearn S, Radecki S. Assessing the acquisition of core clinical skills through the use of serial standardized patient assessments. *Acad Med*. 2000; 75:480-483.
- Jefferies A, Simmons B, Regehr G. The effect of candidate familiarity on examiner OSCE scores. *Med Educ*. Sep 2007;41(9):888-891.
- Regehr G, Freeman R, Robb A, Missiha N, Heisey R. OSCE performance evaluations made by standardized patients: Comparing checklist and global rating scores. *Acad Med*. Oct 1999;74(10 Suppl):S135-S137.
- Solomon DJ, Szauter K, Rosebraugh CJ, Callaway MR. Global ratings of student performance in a standardized patient examination: Is the whole more than the sum of the parts? *Adv Health Sci Educ Theory Pract*. 2000;5(2):131-140.
- Holmboe ES, Yepes M, Williams F, Huot SJ. Feedback and the mini clinical evaluation exercise. *J Gen Intern Med*. May 2004;19(5 Pt 2):558-561.
- Hodder RV, Rivington RN, Calcutt LE, Hart IR. The effectiveness of immediate feedback during the objective structured clinical examination. *Med Educ*. 1989;23(2):184-188.
- Farrell SE. Evaluation of student performance: Clinical and professional performance. *Acad Emerg Med*. 2005;12(4):302e6-302e10.
- Nuovo J, Bertakis KD, Azari R. Assessing resident's knowledge and communication skills using four different evaluation tools. *Med Educ*. 2006;40(7):630-636.

Author Information

Désirée Lie, MD, MEd

Director of Research and Faculty Development, University of California, Irvine School of Medicine

Jennifer Encinas

Ex-research Assistant, University of California, Irvine School of Medicine

Frances Stephens, MFA

Editor and Program Coordinator, University of California, Irvine School of Medicine

Michael Prislin, MD

Associate Dean for Student Affairs, University of California, Irvine School of Medicine