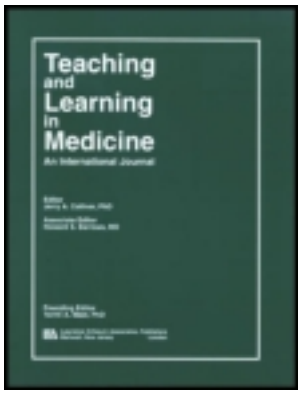

Teaching and Learning in Medicine: An International Journal

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/htlm20

Mobile Technology for the Facilitation of Direct Observation and Assessment of Student Performance

Gary S. Ferenchick [a], David Solomon [b], Jami Foreback [c], Basim Towfiq [d], Kevin Kavanaugh [e], Larry Warbasse [f], James Addison [g], Frances Chames [h], Alvin Dandan [a] & Asad Mohmand [i]

[a] Department of Medicine, Michigan State University College of Human Medicine, East Lansing, Michigan, USA

[b] Division of General Medicine and Office of Medical Education Research and Development, Michigan State University, East Lansing, Michigan, USA

[c] Department of Medicine, Michigan State University College of Human Medicine Flint McLaren Campus, Flint, Michigan, USA

[d] Department of Medicine, Michigan State University College of Human Medicine Flint Hurley Campus, Flint, Michigan, USA

[e] Department of Medicine, Michigan State University College of Human Medicine Kalamazoo Campus, Kalamazoo, Michigan, USA

[f] Department of Internal Medicine, Michigan State University College of Human Medicine Munson Campus, Traverse City, Michigan, USA

[g] Department of Internal Medicine, Michigan State University College of Human Medicine Marquette General Campus, Marquette, Michigan, USA

[h] Department of Medicine, Michigan State University College of Human Medicine Grand Rapids Campus, Grand Rapids, Michigan, USA

[i] Virginia Commonwealth University, Richmond, Virginia, USA

Published online: 10 Oct 2013.

To cite this article: Gary S. Ferenchick, David Solomon, Jami Foreback, Basim Towfiq, Kevin Kavanaugh, Larry Warbasse, James Addison, Frances Chames, Alvin Dandan & Asad Mohmand (2013) Mobile Technology for the Facilitation of Direct Observation and Assessment of Student Performance, Teaching and Learning in Medicine: An International Journal, 25:4, 292-299

To link to this article: http://dx.doi.org/10.1080/10401334.2013.827972

Teaching and Learning in Medicine, 25(4), 292–299

ISSN: 1040-1334 print / 1532-8015 online
DOI: 10.1080/10401334.2013.827972

Routledge
Taylor & Francis Group

Mobile Technology for the Facilitation of Direct Observation and Assessment of Student Performance

Gary S. Ferenchick
Department of Medicine, Michigan State University College of Human Medicine, East Lansing, Michigan, USA

David Solomon
Division of General Medicine and Office of Medical Education Research and Development, Michigan State University, East Lansing, Michigan, USA

Jami Foreback
Department of Medicine, Michigan State University College of Human Medicine Flint McLaren Campus, Flint, Michigan, USA

Basim Towfiq
Department of Medicine, Michigan State University College of Human Medicine Flint Hurley Campus, Flint, Michigan, USA

Kevin Kavanaugh
Department of Medicine, Michigan State University College of Human Medicine Kalamazoo Campus, Kalamazoo, Michigan, USA

Larry Warbasse
Department of Internal Medicine, Michigan State University College of Human Medicine Munson Campus, Traverse City, Michigan, USA

James Addison
Department of Internal Medicine, Michigan State University College of Human Medicine Marquette General Campus, Marquette, Michigan, USA

Frances Chames
Department of Medicine, Michigan State University College of Human Medicine Grand Rapids Campus, Grand Rapids, Michigan, USA

Alvin Dandan
Department of Medicine, Michigan State University College of Human Medicine, East Lansing, Michigan, USA

Asad Mohmand
Virginia Commonwealth University, Richmond, Virginia, USA

Background: **We developed, implemented, and assessed a web-based clinical evaluation application (i.e., CEX app) for Internet-enabled mobile devices, including mobile phones. The app displays problem-specific checklists that correspond to training problems created by the Clerkship Directors in Internal Medicine.** ***Purpose:*** **We hypothesized that use of the CEX app for directly observing students' clinical skills would be feasible and acceptable, and would demonstrate adequate reliability and validity.** ***Methods:*** **Between July 2010 and February 2012, 266 third-year medical students completed 5 to 10 formative CEXs during their internal medicine clerkship. The observers (attendings and residents), who performed the CEX, used the app to guide and document their observations, record their time observing and giving feedback to the students, and their overall satisfaction with the CEX app. Interrater reliability and validity were assessed with 17 observers**

This project was funded in part by the Health Resources and Services Administration Grant D56HP05217. We thank Anton Frattaroli, Josh Frank and Steve Sneed, who provided the programming support for the CEX app.

Correspondence may be sent to Gary Ferenchick, Michigan State University, Clinical Center, 788 Service Road, B335, East Lansing, MI 48824, USA. E-mail: gary.ferenchick@hc.msu.edu

who viewed 6 videotaped student–patient encounters, and by measuring the correlation between student CEX scores and their scores on subsequent standardized-patient Objective Structured Clinical Examination (OSCE) exams. *Results:* A total of 2,523 CEXs were completed by 411 observers. The average number of evaluations per student was 9.8 (± 1.8 *SD*), and the average number of CEXs completed per observer was 6 (± 11.8 *SD*). Observers spent less than 10 min on 45.3% of the CEXs and 68.6% of the feedback sessions. An overwhelming majority of observers (90.6%) reported satisfaction with the CEX. Interrater reliability was measured at 0.69 among the observers viewing the videotapes, and their ratings discriminated between competent and noncompetent performances. Student CEX grades, however, did not correlate with their end of 3rd-year OSCE scores. *Conclusions:* The use of this CEX app is feasible and it captures students' clinical performance data with a high rate of user satisfaction. Our embedded checklists had adequate interrater reliability and concurrent validity. The grades measured on this app, however, were not predictive of subsequent student performance.

Keywords medical education, educational assessment, clinical clerkship, clinical competence, observation, computers, smartphones, checklists, cloud application, patient encounters

INTRODUCTION

Direct observation of students' performance in clinical settings remains challenging. The rationale for direct observation is to ascertain and document that students have acquired core clinical skills. Direct observation has been shown to improve the interrater reliability and predictive validity of other evaluative measures when used as an additional source of acquiring clinical performance data.[1] Such observation, however, occurs infrequently, and when it does occur, it produces evaluations that can be highly variable and skewed by factors other than a trainee's actual performance.[2–4]

There are a number of barriers to direct observation including the lack of effective implementation strategies and the lack of transparent standards used in judging clinical competence.[1,5] Studies have demonstrated that in the absence of explicit criteria, the discriminating ability of evaluators judging the exact same clinical performance is poor.[4,6] The development of external standards, such as checklists for specific clinical problems, for use by observers during student–patient encounters offers a potential solution to this problem. Experts in the field have advocated for such criterion-based assessment.[1,4]

One of the few feasible ways to efficiently distribute problem-specific criterion-based checklists to observers at the point of assessment is with mobile technology. We have previously demonstrated the feasibility of displaying such competency checklists on electronic handheld personal digital assistants (PDAs) such as Palm® and PocketPC® devices.[7] In that study, the use of checklists displayed on PDAs (i.e., eCEXs) facilitated the direct observation and assessment of students in our geographically dispersed medical school.[7] Students and preceptors valued the fact that the assessed competencies were transparent. However, technical issues with the PDAs—such as lack of direct Internet connection and the requirement for students to "synchronize" data from the PDA to the web using a desktop computer—limited their overall usefulness; a process that is not needed with contemporary Internet-enabled mobile devices such as iPads, iPhones, and other smartphones. These mobile devices have become almost ubiquitous since the time of our last study, and we leveraged this trend to evolve our eCEX to a web-based and platform-neutral clinical evaluation application we are calling the "CEX app."

There are three major intended utilities of the CEX app: (a) to facilitate the direct observation of students' clinical skills during encounters with real patients, (b) to increase the transparency of the expected competencies for these observations for both the observers and the student, and (c) to have a permanent record of these observed skills.

For our current study we hypothesized that use of the CEX app would be feasible, be generally accepted by faculty and resident raters, and have adequate interrater reliability and validity.

METHOD

Description of the Technology

A web-based content management system is the backbone of our CEX app. We have previously published the details of this system.[7,8] Briefly, this system allows users to create customizable content including problem-specific checklists that can be displayed on Internet-enabled mobile devices. Tasks performed by the students can be directly marked as "completed" on these checklists by faculty and/or resident observers. We developed customized checklists for our internal medicine clerkship focusing on student-level competencies for communication, history-taking, and physical examination skills pertaining to 16 core training problems described by the Clerkship Directors in Internal Medicine (CDIM). The core CDIM curriculum was developed and validated in 1996 and subsequently revised in 1998 and 2006. The curriculum is organized around specific training problems common to internal medicine. The CDIM training problems have been used extensively in medicine clerkships throughout the United States.[9,10]

In addition to these 16 problems, checklists for the problem "Headache" (not a core CDIM problem), and three separate counseling skills (informed decision making, smoking cessation, and substance abuse history taking) were also included. The same set of clinical problems is used for our mandated patient encounter log.[11] Screenshots of the CEX app and dashboard of completed student CEXs are shown in Figures 1 and 2.*

*The CEX app can be accessed with any Internet connection at
http://www.justintimemedicine.com/mobile
Username: testuser@tlm.edu
PW: test
Cloud-based reports can be viewed at www.justintimemedicine.com; log in is testadministrator@tlm.edu and the PW: test; click assessment reports (UME-Style)

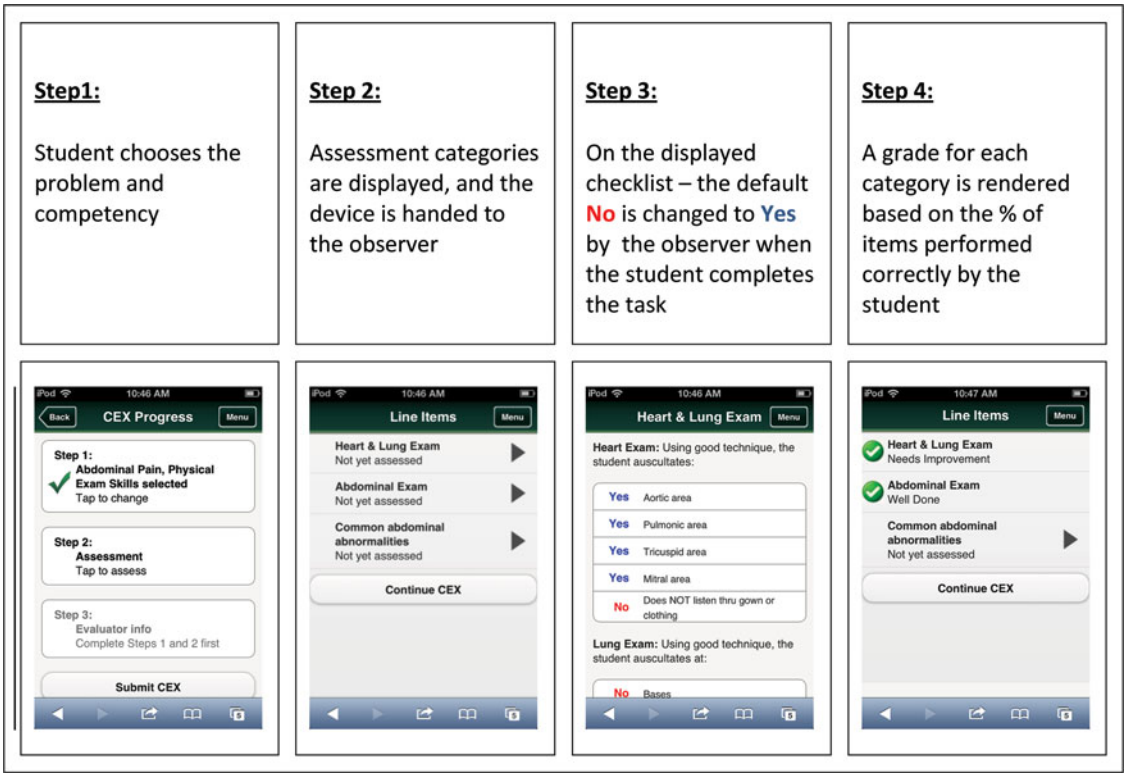

FIG. 1. Screenshots of performance checklists as displayed on an iPhone (color figure available online).

Setting

Between July 2010 and February 2012, 266 third-year medical students at Michigan State University's College of Human Medicine (CHM) completed an 8-week core internal medicine clerkship in seven geographically dispersed communities throughout the state. As a requirement for passing the clerkship, all 21 students in two of our seven communities were required to complete five CEX evaluations and a full history

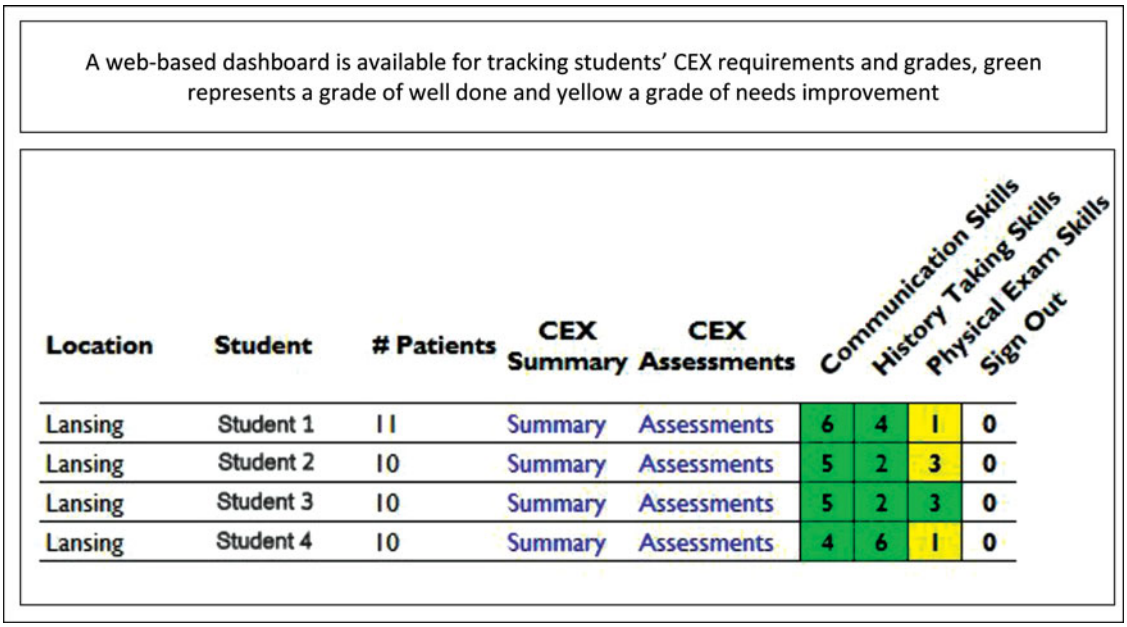

FIG. 2. Screenshot of a cloud-based dashboard tracking student clinical evaluation (CEX) performance, with the capability to drill down to more detailed reports (color figure available online).

and physical examination while being observed; in the other five communities all 245 students were required to complete 10 CEX evaluations with their clinical preceptors. All CEX evaluations were for formative assessment. For each evaluation, the individual student would choose the problem (e.g., abdominal pain), the competency (e.g., history taking), and the specific observer (an attending or resident physician).

At the beginning of each 8-week rotation, students were e-mailed information on the CEX app, which included links to online tutorials, and they were given a 15-min live overview on the CEX. They were encouraged to familiarize themselves with the screen flows with several practice entries. Students were also responsible for orienting and assisting the observer, as needed, on the use of the CEX app. In addition, all evaluators had access to an online tutorial on the CEX app.

Students could choose from a menu of 51 possible options to meet the required number of CEX evaluations (Table 1). For example, for 14 of the core training problems, a student could choose to be assessed on communication, history-taking skills, or physical exam skills (42 potential options). For three of the core training problems (depression, dysuria, and fluid & electrolyte abnormalities) only communication skills or history taking could be selected (six potential options), and for informed decision making, smoking cessation, and substance use disorders, only a single competency could be chosen.

TABLE 1
Menu of CEX options from which students can choose

Training Problem	Communication Skills	History-Taking Skills	Physical Exam Skills
Abdominal Pain	16	47	25
Altered Mental Status	17	31	37
Anemia	16	47	24
Back Pain	23	36	17
Congestive Heart Failure	17	55	23
COPD/Asthma	14	37	19
Chest Pain	17	48	28
Cough	17	31	16
Depression	16	43	
Diabetes	17	32	19
Dysuria	17	22	
Fluid and Electrolytes	17	21	
GI Bleeding	17	43	26
Headache	17	44	29
Hypertension	17	47	30
Joint Pain	17	44	39
Liver Disease	17	35	23
Informed Decision Making	6		
Smoking Cessation	15		
Substance Use Disorders	18		

Note. Gray cells represent the 51 clinical evaluation (CEX) options for student assessment. The number of checklist items is listed for each problem and competency. For example, there are 25 separate items to assess for the training problem "abdominal pain" and the competency "physical exam". White cells represent no CEX option for the student. COPD = chronic obstructive pulmonary disease; GI = gastrointestinal.

Each checklist contained multiple specific items that a student was expected to address for a given condition. The checklist scales were criterion based (e.g., "Student specifically asks about, elicits, or establishes the presence or absence of weight change" Yes/No). Students were encouraged to view the performance checklists prior to each assessment. Table 2 displays an example of the items listed for history taking in a patient with abdominal pain.

Using checklists displayed on a students Internet-enabled device, observers documented the student's performance but did not "grade" or render a judgment on the student's performance per se. Whereas the CEX sessions remained formative, they were graded automatically by an algorithm built into the CEX app which equally weighted all items. A grade of "well done" was rendered if a student completed a prespecified percentage of the items on the checklists; a grade of "needs improvement" was rendered for a score less than this prespecified percentage, generally 70% to 80%; these percentage correct scores correspond to the number of items required to pass our end-of-year Objective Structured Clinical Examination (OSCE). Observers recorded the amount of time they spent during the student–patient encounter, the amount of time they spent giving feedback, and their overall satisfaction with the CEX on a 5-point Likert scale after each completed session. The checklist data were rendered complete only after the observer clicked the "Submit CEX" button on the app, which then stored the results of the evaluation in an online database (Figure 2).

Data Collection

We collected and analyzed all CEX evaluations. The data included specific problems and observed competencies, the percentage of CEXs with a "well done" grade, the average number of checklist items completed correctly, and descriptive data from observers on the use of the CEX app. We ran the descriptive statistics separately for resident observers and attending observers.

Validity and Reliability Assessments

A group of 17 observers (nine internal medicine residents and eight general internal medicine faculty) viewed and rated six scripted videotaped encounters. There were two replicates each of a standardized trainee taking a history from a standardized patient with chest pain, performing a physical exam on a patient with dyspnea and engaging in a counseling session with a patient starting a new medication. For each replicate, one encounter was scripted as clearly demonstrating an unsatisfactory performance (less skill and many deficiencies) and the other was scripted as a clearly satisfactory performance (moderate to excellent skill and fewer deficiencies). These cases have been previously used

TABLE 2
Example of a problem-specific checklist: History taking in a patient with abdominal pain

Doctor–Patient Communication	Associated GI Symptoms
Student specifically: • Asks how the pain has affected the patient's life • Asks how the patient is coping • Demonstrates concern for the patient • Responds with at least 1 statement of empathy	Student asks directly about: • Diarrhea • Constipation • Bright red blood in stool • Melena • Hematemesis • Vomiting
Key Pain Characteristics	**Past Medical, Family and Social History**
Student specifically asks about or establishes: • Location • Character • Intensity (scale 1–10) • Onset: sudden vs. gradual • Pattern: acute vs. chronic • Pattern: constant vs. intermittent • Pattern: accelerating or not • Radiation • Duration of each episode • Total duration of the problem • Aggravating/alleviating factors	Past history: Student specifically asks about, elicits or establishes the presence OR ABSENCE of: • PUD • Pancreatitis • Biliary disease • Liver disease • Renal stones • Diverticular disease • GI surgery • Pelvic surgery Family history: Student establishes presence OR ABSENCE of: • Family history of GI disease Social history: Specifically asks about, elicits or establishes presence OR ABSENCE of: • Alcohol use • Illicit drug use • Tobacco use • Sexual activity (risk of STD) • Psychosocial stressors • Financial concerns for care
Associated Symptoms	**Medications**
Other symptoms: Student specifically asks about, elicits or establishes presence OR ABSENCE of: • Weight change • Fever • Dysuria	Student specifically asks about, elicits or establishes: • All medications, including OTC • Supplements/herbs • Specifically asks about NSAID use (including OTC) • If diarrhea present, asks about recent antibiotic use

Note. PUD = peptic ulcer disease; GI = gastrointestinal; STD = sexually transmitted disease; OTC = over the counter.

for faculty development and validated as representing levels of competence that range from unequivocally poor to satisfactory.[6]

CHM conducts a high-stakes OSCE-like performance assessment examination at the end of the 3rd year. It consists of 10 cases, of which six are graded. The exam is focused on communication, history taking, and physical examination skills. For the purpose of assessing the predictive validity of the CEX grades, we correlated the summative OSCE scores for communication, history taking, and physical examination for 150 third-year students with their CEX grades for the corresponding competency.

Human Subject Protection

Our medical school has created an "Honest Broker System" for conducting research on student performance data that are collected as a regular part of the students' educational activities. A designated employee of the medical school, with access to

these data, has been designated as the Honest Broker. In an honest broker system a person or agency that has access to multiple human subject datasets collected for nonresearch purposes creates a deidentified dataset that can be used for research purposes without posing risk to the subjects.[12] This approach has been used in various types of clinical research—though, to the best of our knowledge, it has not been applied in educational research at other institutions.

At Michigan State University this designated individual created the dataset used in this study and made it available to our research team after removing all identifiers. The Social Science/Behavioral/Education Institutional Review Board at Michigan State University granted our study an exempted review.

RESULTS

Number and Types of Evaluations

Four hundred eleven observers (attending and resident physicians) completed 2,523 CEXs with 266 students. The average number of CEX evaluations per student was 9.8 ($SD = 1.8$), and the average number of CEXs completed per attending or resident observers was six ($SD = 11.8$). Students were evaluated on an average of seven of a possible 20 problems, the most frequent ones being abdominal pain, altered mental status, chest pain, congestive heart failure, chronic obstructive pulmonary disease, dyspnea, and headache.

Of the three competency domains of communication skills, history taking, and physical examination, 71.8% of students were evaluated on all three, 23.3% on two, and 4.9% on a single competency.

Grading

A "well done" grade was received for 86% of the CEX evaluations and a "needs improvement" grade was received for 14% of the CEX evaluations. Quantitatively, students averaged 86.4% (±16.2) for all the potential items correct, with a high of 90.1% correct for congestive heart failure and a low of 82.7% correct for abdominal pain.

Observer Time and Satisfaction

Observers reported that 45.3% of the CEXs took them less than 10 min to complete and that feedback lasted less than 10 min for 68.6% of the CEXs. We compared CEXs from the 90 attending physicians and 314 residents who observed students over the study period.* We found no meaningful differences between resident and faculty observers in reported time for observation and feedback. Residents were slightly more positive about rating the CEX experience, although satisfaction was quite high across both groups. Fifty-seven percent of the faculty attendings were satisfied with the CEX experience in comparison with 44% of the residents, whereas 48% of the residents were very satisfied as compared with 32% of the faculty (Table 3).

TABLE 3
Descriptive data on the use of the CEX app by faculty and resident evaluators

Time Spent Completing the CEX (Reported by the Observer)			
Time Reported to Complete the CEX	Faculty Attending (%)	Resident (%)	Total
1–5 Min	8.1	8.7	8.5
6–10 Min	40.7	34.7	36.8
11–15 Min	22.1	29.8	27.1
>15 Min	29.1	26.9	27.6
Time Spent by Observers in Providing Feedback to Students			
1–5 Min	32.1	33.6	33.0
6–10 Min	35.9	35.5	35.6
11–15 Min	17.1	17.2	17.2
>15 Min	14.9	13.7	14.1
Observers Satisfaction With the CEX			
Highly Satisfied	30.9	48.1	42.1
Satisfied	57.3	43.8	48.5
Neutral	9.5	7.0	7.8
Dissatisfied	1.5	0.4	0.8
Highly Dissatisfied	0.8	0.7	0.7

Note. CEX = clinical evaluation.

Residents tended to rate history taking more often than faculty attendings (42% vs. 26%), whereas faculty attendings rated physical exam skills more often than the residents (60% vs. 47%).

Validity and Reliability

The mean interrater reliability was 0.69 for the observers viewing the six videotaped encounters. The interrater reliability rating was 0.74 for general internal medicine faculty and 0.64 for internal medicine residents. Observer ratings discriminated between the encounters scripted for unsatisfactory performance and those scripted for satisfactory performance. For example, 35% to 59% fewer items were rated as being performed correctly on the nonsatisfactory encounters compared with the satisfactory encounters. This difference was statistically significant ($p < .0001$) based on a matched pairs t test. There was no statistically significant correlation between CEX grades and the corresponding scores on the end-of-year OSCE.

CONCLUSIONS

Internet-enabled mobile devices such as smartphones and tablets are becoming ubiquitous. We leveraged this trend to develop, implement, and assess a web-based clinical evaluation application for such devices. This application, the CEX app, displays customized problem-specific checklists that display specific performance standards to use as a reference for observers

*Differences in numbers due to missing data for seven evaluators.

and evaluators at the time of a planned direct-observation. We have previously demonstrated that this feature fosters a high degree of understanding among faculty and students of what to expect during the evaluation. Faculty also reported that this feature improved their assessments and ability to give feedback.[7] Our current study extends these findings in that the use of this application was associated with high user satisfaction among 411 observers in our geographically dispersed medical school. It also shows that the checklists displayed on the CEX app for history taking in a patient with chest pain, physical examination in a patient with dyspnea, and informed decision making possess adequate interrater reliability and concurrent validity.

The number of times students report being directly observed by faculty during patient encounters continues to be unacceptably low.[2,3] In the 2011 AAMC graduation survey, up to 36% of graduating students disagreed with the statement "A faculty member personally observed me taking a patient history (or performing physical examinations) during the clerkship."[13] Previous studies focusing on student-initiated direct observation have reported varying degrees of success in assuring that the observation actually occurred. In one study, observations occurred 0.5 times per student per week, falling short of the stated goal of two per student per week.[14] In other studies the actual accomplishment of the CEXs was 13 to 59% lower than the stated goal for the number of CEX's expected.[5,15] Strategies and interventions are needed to assure that all students are observed by faculty, and our work in this regard has demonstrated that mobile technology facilitates the completion of direct observations.[7,15,16]

It is important to familiarize faculty observers with specific performance standards when observing and assessing a learner's clinical competencies. As many authors have noted, when direct observation does occur, faculty observers are commonly unsure of the standards for judging the performance of learners.[1,5,17] Given this uncertainty, faculty observers tend to view a trainee's clinical skills through a myriad of lenses including other practicing doctors, themselves, gestalt, the use of unsubstantiated inferences, and so on.[4] This results in marked variability and subjectivity in subsequent assessments and evaluations. Observers often fail to use existing frameworks or external standards to guide them in evaluating the performance of a trainee.[4] Thus, it is not surprising that when faculty members view the exact same videotaped trainee–patient encounters, they commonly render different evaluations and interpretations.[4,6]

In this study, we instructed observers to monitor and document a student's performance but to not "grade" the student. Kogan has demonstrated that most faculty members have difficulty articulating how they move from observation to judgment and from judgment to assigning a numerical grade.[4] In our study, we implemented a process where the grade was electronically rendered based on a simple algorithm that reflected the percentage of expected problem-specific tasks performed correctly by the students. Thereby, we took the observer out of the role of judging and rating the students' performances. Anecdotally, several faculty attendings and residents participating in this process appreciated not having to render a grade. We, therefore, feel that the use of electronic algorithms that automatically assign a "grade" to student performance based on an objective assessment of what the student does and does not accomplish in a clinical encounter deserves further study, as does the development of algorithms that weigh items related to their importance in demonstrating basic competency.

Fromme noted in a literature review in 2009 that most studies on observation do not specify how patients were selected for direct observation. She noted that patients were commonly chosen on the basis of convenience, and not on the basis of educational goals.[18] We addressed this concern by using the same list of the types of patients and clinical conditions a student must encounter for the CEX observations as we have for our patient logs. This has enabled us to move closer to attaining congruence between nationally accepted core content, the patient log expectations for our students, and our assessment methods. The importance of this, and our grading incentive (i.e., a grade is given only if students complete the specified number of CEX assessments), is that students are motivated to choose specific patients for the CEX assessment for educational reasons and not simply on the basis of convenience.

According to Fromme, content validity represents the degree to which "learners are being tested on what is intended to be taught."[18] We adopted most of our checklists from problems specified in the CDIM core curriculum guide in order to assure the validity of our checklists. We also demonstrated concurrent validity in that observer ratings of the videotaped encounters appropriately separated unsatisfactory from satisfactory performances. We established adequate interrater reliability, with faculty attending raters showing slightly higher reliability compared with resident raters. However, given the binary nature of the observations (the student either performed the required skill or not), one might have expected a higher interrater reliability than what we actually achieved. All checklists, including ours, have fundamental conflicts between effectiveness and brevity. Consequently the number of screens required to display our checklists, which averaged 16 discrete items for each physical examination CEX and 30 discrete items for each history taking CEX, may have influenced our interrater reliability. Future studies need to assess not only the minimum number of items required per skill set to determine competence but also the maximum number of screen clicks an assessor can reasonably accomplish in a 10-min observation.

We did not, however, establish the predictive validity of scores obtained on the CEX app. There are several possible explanations for this. First, between 1 and 11 months separated the CEX and OSCE results, during which time students continued in required 3rd-year clerkships. Second, our CEX assessments did not precisely mirror our OSCE assessments, which sample clinical skills from across many disciplines (including pediatrics, obstetrics, psychiatry, and surgery) and our students could review the CEX checklists prior to the planned assessment, an option not available for the OSCE assessments. Third, the students were given the freedom to choose their assessors in

the CEX, an option not offered in the OSCE, and this could have biased the assessments. Finally our OSCE was completed with standardized patients in controlled settings, whereas our CEXs were completed with real patients in clinic settings associated with complex variables that can, in theory, influence student performance (time factors, the acuity of the patient's illness, the presence of family members, etc.). Future studies should assess the predictive value of the CEX performance related to OSCE performance in a more nuanced way, such as assessing the CEX performance against only the OCSE cases related to internal medicine, and comparing CEX performances within a few months of the OSCE to those that were completed at more remote times (e.g., 6–10 months).

Programs have previously been developed for PDAs for use in clinical and medical education.[7,8,11,16] Studies demonstrate that handheld device-based evaluations make the logistics of CEX observations easier. Torre demonstrated that acceptability and reliability of PDA and paper-based mini-CEX were comparable. However, clinical competency scores and observation and feedback times were higher in the PDA-based mini-CEX assessments compared with the paper-based counterpart.[16]

The efficiency of the PDA-based mini-CEX is reproducible in today's smartphones. A recent publication described the development of an iPad platform for the mini-CEX. However, this iPad mini-CEX application was tested by only five subjects and its utility in the actual clinical setting is yet to be determined.[19]

Our study, like most others, has several inherent limitations. First, this is a single institution study and these results may not be generalizable. Future studies should focus on the use of this technology in more diverse settings. Second, establishing the reliability of all of the customized checklists within the CEX app is needed, as is establishing its reliability in real clinical settings such as the hospital wards. Third, the electronic grading algorithm employed by the CEX app is not yet systematically validated. Last, like many tools for direct observation, we have not established the effect of this tool on learning, transfer of acquired clinical skills to other areas, or the effect that such direct observation has on patient care.

We conclude that the use of smartphones and other Internet-enabled mobile devices makes the effective use of the CEX app feasible. Software developed for these devices simplifies the delivery and dissemination of customizable checklists for use by clinical educators in the real-world assessment of their trainees' clinical skills. More than 400 observers, including faculty and residents, used our CEX app with a high degree of satisfaction, making it a practical and innovative tool in the otherwise busy routines of most clinical educators today. Although not specifically addressed in this study, we believe that creating customized checklists displayable for use on smartphones has great potential in meeting many of the goals of competency-based medical education, including documenting entrustable professional activities (EPAs).

REFERENCES

1. Hasnain M, Connell KJ, Downing SM, Olthoff A, Yudkowsky R. Toward meaningful evaluation of clinical competence: the role of direct observation in clerkship ratings. *Academic Medicine* 2004;79:S21–4.
2. Pulito AR, Donnelly MB, Plymale M, Mentzer RM Jr. What do faculty observe of medical students' clinical performance? *Teaching and Learning in Medicine* 2006;18:99–104.
3. Howley LD, Wilson WG. Direct observation of students during clerkship rotations: a multiyear descriptive study. *Academic Medicine* 2004;79: 276–80.
4. Kogan JR, Conforti L, Bernabeo E, Iobst W, Holmboe E. Opening the black box of clinical skills assessment via observation: A conceptual model. *Medical Education* 2011;45:1048–60.
5. Fernando N, Cleland J, McKenzie H, Cassar K. Identifying the factors that determine feedback given to undergraduate medical students following formative mini-CEX assessments. *Medical Education* 2008;42: 89–95.
6. Holmboe ES, Hawkins RE, Huot SJ. Effects of training in direct observation of medical residents' clinical competence: A randomized trial. *Annals of Internal Medicine* 2004;140:874–81.
7. Ferenchick GS, Foreback J, Towfiq B, Kavanaugh K, Solomon D, Mohmand A. The implementation of a mobile problem-specific electronic CEX for assessing directly observed student-patient encounters. *Medical Education Online* 2010;15;4276. doi:10.3402/meo.v/510.4276
8. Ferenchick G, Fetters M, Carse AM. Just in time: Technology to disseminate curriculum and manage educational requirements with mobile technology. *Teaching and Learning in Medicine* 2008;20:44–52.
9. Denton GD, Durning SJ. Internal medicine core clerkships experience with core problem lists: Results from a national survey of clerkship directors in internal medicine. *Teaching and Learning in Medicine* 2009;21: 281–3.
10. Jablonover RS, Blackman DJ, Bass EB, Morrison G, Goroll AH. Evaluation of a national curriculum reform effort for the medicine core clerkship. *Journal of General Internal Medicine* 2000;15:484–91.
11. Ferenchick G, Mohmand A, Mireles J, Solomon D. Using patient encounter logs for mandated clinical encounters in an internal medicine clerkship. *Teaching and Learning in Medicine* 2009;21:299–304.
12. Boyd AD, Hosner C, Hunscher DA, Athey BD, Clauw DJ, Green LA. An 'Honest Broker' mechanism to maintain privacy for patient care and academic medical research. *International Journal of Medical Informatics* 2007;76:407–11.
13. 2011 AAMC Graduation Questionairre Program Evaluation Survey: All Schools Summary Report Final. https://www.aamc.org/data/gq/allschoolsreports. Accessed April 23, 2012.
14. Kuo AK, Irby DI, Loeser H. Does direct observation improve medical students' clerkship experiences? *Medical Education* 2005;39:518.
15. Kogan JR, Bellini LM, Shea JA. Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship. *Academic Medicine* 2003;78;10:S33–S5.
16. 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. *Teaching and Learning in Medicine* 2007;19:271–7.
17. Hill F, Kendall K, Galbraith K, Crossley J. Implementing the undergraduate mini-CEX: A tailored approach at Southampton University. *Medical Education* 2009;43:326–34.
18. Fromme HB, Karani R, Downing SM. Direct observation in medical education: A review of the literature and evidence for validity. *Mt Siani Journal of Medicine New York* 2009;76:365–71.
19. DS. Development and beta testing of applications for direct trainee observation in medical education using an iPad platform. *The Research Bulletin of Jordan ACM - ISWSA*. 2011;11:35–40.