

Using systematically observed clinical encounters (SOCEs) to assess medical students' skills in clinical settings

George R Bergus¹⁻³
Jerold C Woodhead⁴
Clarence D Kreiter^{2,5}

¹Performance Based Assessment Program, Office of Student Affairs and Curriculum, ²Department of Family Medicine, ³Department of Psychiatry, ⁴Department of Pediatrics, ⁵Office of Consultation and Research in Medical Education, Roy J and Lucille A Carver College of Medicine, The University of Iowa, Iowa City, IA, USA

Introduction: The Objective Structured Clinical Examination (OSCE) is widely used to assess the clinical performance of medical students. However, concerns related to cost, availability, and validity, have led educators to investigate alternatives to the OSCE. Some alternatives involve assessing students while they provide care to patients – the mini-CEX (mini-Clinical Evaluation Exercise) and the Long Case are examples. We investigated the psychometrics of systematically observed clinical encounters (SOCEs), in which physicians are supplemented by lay trained observers, as a means of assessing the clinical performances of medical students.

Methods: During the pediatrics clerkship at the University of Iowa, trained lay observers assessed the communication skills of third-year medical students using a communication checklist while the students interviewed and examined pediatric patients. Students then verbally presented their findings to faculty, who assessed students' clinical skills using a standardized form. The reliability of the combined communication and clinical skills scores was calculated using generalizability theory.

Results: Fifty-one medical students completed 199 observed patient encounters. The mean combined clinical and communication skills score (out of a maximum 45 points) was 40.8 (standard deviation 3.3). The calculated reliability of the SOCE scores, using generalizability theory, from 10 observed patient encounters was 0.81. Students reported receiving helpful feedback from faculty after 97% of their observed clinical encounters.

Conclusion: The SOCE can reliably assess the clinical performances of third-year medical students on their pediatrics clerkship. The SOCE is an attractive addition to the other methods utilizing real patient encounters for assessing the skills of learners.

Keywords: performance assessment, clinical skills, medical education

Introduction

Clinical competence requires mastery and integration of specific skills (history taking, physical examination, and communication skills), knowledge, and clinical reasoning.¹ Traditional methods of assessing the clinical performance of learners, based on brief casual faculty observations, have limited validity.² In 1979 Harden published his innovative work on utilizing the Objective Structured Clinical Examination (OSCE) for assessing learners.³ During an OSCE, simulated patients using explicit scripts are interviewed and examined by students while the students' actions are recorded by observers. This standardization of simulated clinical encounters and the rating instruments markedly improved reliability as compared with nonstandardized real patient encounters.

Although most US medical schools currently use the OSCE to assess their students' skills, further development of performance-based skills assessment is desirable.^{4,5}

Correspondence: George R Bergus
Department of Family Medicine,
200 Hawkins Drive, The University
of Iowa, Iowa City, IA 52242, USA
Tel +1 319 384 7566
Fax +1 319 384 7822
Email george-bergus@uiowa.edu

For example, infants and young children are frequently encountered in clinical settings but are difficult to incorporate into OSCEs. The same is true for elderly and frail patients. In addition, the OSCE's separation from the clinical environment has also caused some medical educators to suggest that the OSCE may adversely affect how future doctors learn to interact with patients.^{6,7}

Research suggests that reliable performance assessments of students can be undertaken in clinical settings. Educators have argued that these assessment tools have advantages compared with the OSCE.⁸⁻¹⁰ One approach, the mini-Clinical Evaluation Exercise (mini-CEX), has faculty observe students during an encounter with a patient.^{11,12} A second approach is to have students interview and examine patients and then make a verbal presentation of the findings to the faculty. This method, known as the Long Case, results in scores with a reliability similar to the OSCE if the students' interactions with the patients are directly observed by the faculty.^{13,14} As a potential alternative to the Long Case and the mini-CEX, we developed the systematically observed clinical encounter (SOCE). The SOCE is similar to the Long Case in that students are assessed during encounters with actual patients, and faculty rate the clinical skills of students based on their presentations. However, unlike the Long Case, students are not observed by faculty but by standardized observers.

We previously reported that standardized observers can reliably score the communication skills of students by observing the students' clinical interactions with real patients.¹⁵ We now report on the SOCE as a comprehensive measure of medical students' clinical performance resulting from combining the standardized observers' communication skills scores with clinical skills ratings from faculty. In addition, we report on student perceptions about the SOCE as a method of evaluation.

Methods

Subjects

Third-year medical students at the University of Iowa Carver College of Medicine on a 6-week required pediatrics clerkship were invited to participate in this trial. Participation in this research using the SOCE to assess clinical performance was entirely voluntary and did not affect the grades of students on this clerkship.

Students participating in the trial were assigned pediatric patients per the usual protocol of the General Pediatrics Clinic for students on the rotation. During this clerkship, medical students typically spend a total of 2-4 half-days in this outpatient clinic. This clinic is a primary care medical

office where residents and students are routinely involved in the care of pediatric patients, most of whom have acute medical complaints.

The faculty physicians assessing the clinical skills of the medical students were full-time faculty in the Department of Pediatrics at the University of Iowa and were regularly assigned to oversee patient care and teach medical learners in the General Pediatrics Clinic. The standardized observers who observed the clinical interactions of participating students with their patients had extensive experience with the evaluation of medical students' communication skills during OSCEs. Our standardized observers were lay people without training in the health sciences.

This project was approved by the University of Iowa Institutional Review Board. Funding for this project was obtained from an internal educational development grant from the medical school. The funding body did not have access to the data and was not involved in the analysis of the data or the writing up of the results.

Study protocol

At the start of each half-day clinic session, a standardized observer met with the third-year medical student assigned to the clinic. If the student consented to participate, the standardized observer went into the clinic waiting room to find the patient the student was scheduled to see next. If the patient and his/her parent consented, the standardized observer then accompanied the patient and his/her parent into an exam room. When the student entered the exam room the standardized observer took a position outside of the student's line of sight and completed the communication checklist while observing the student. The standardized observer remained in the exam room for the entire encounter (Figure 1).

After completing his or her evaluation of the patient, the student left the exam room and presented the findings to a faculty physician staffing the clinic as is the routine for students on this clerkship. The student and faculty then discussed the patient and returned to the exam room where the faculty confirmed the findings. Afterwards, the faculty rated the performance of the learner using a 6-item clinical skills rating instrument (see Appendix 1). Four of the items focused on the history the student collected (completeness of the history of the present illness [HPI], development of the HPI in terms of time course and changes, completeness of other appropriate history, and accuracy of the history the student collected). One of the items focused on the completeness of the physical exam completed by the student, and the final item addressed the soundness of the student's clinical

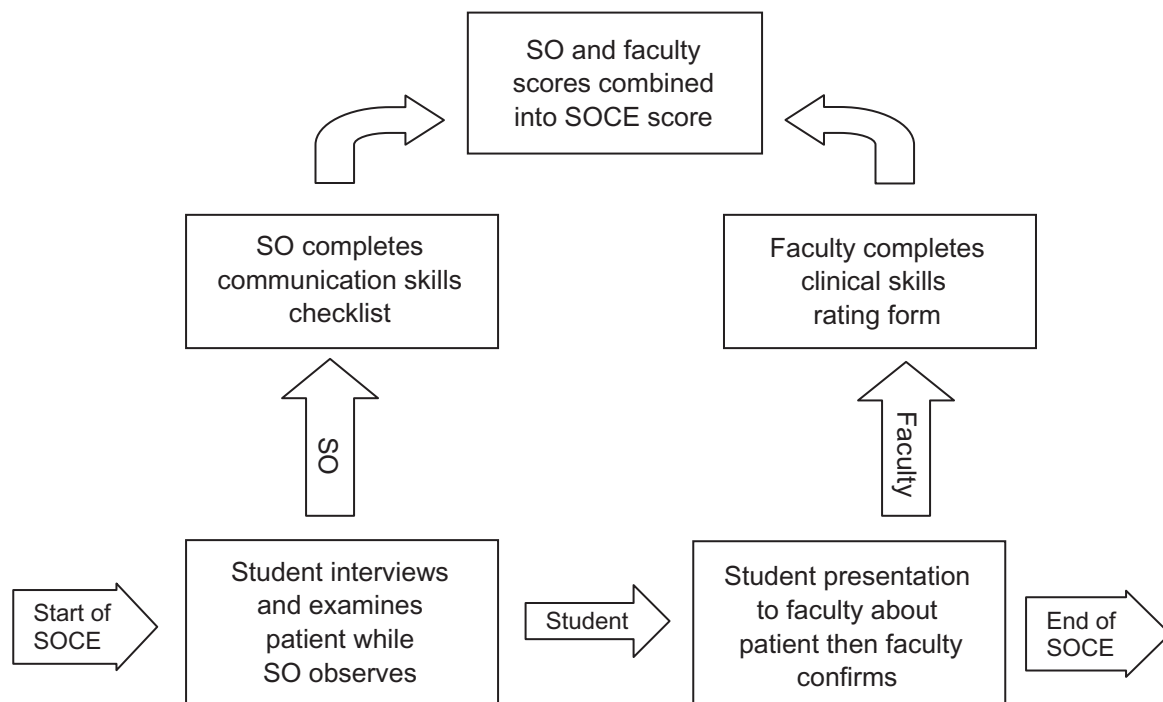


Figure 1 A schematic of the two-step systematically observed clinical encounter (SOCE) process. The student first interviews and examines a patient while being observed by a standardized observer (SO). The student then presents their findings to a faculty physician who is supervising learners in the General Pediatrics Clinic.

reasoning. Thus, by the end of the SOCE, the student had interviewed and examined a patient, and the student's data collection skills, clinical reasoning, and communication skills were assessed by a standardized observer and a faculty physician. At the end of an observed encounter, participating students completed a questionnaire about how the student felt about the experience. This questionnaire contained a series of statements to which students were asked about their level of agreement on a five-point Likert response scale ranging from "strongly agree" to "strongly disagree".

Analysis

Each patient encounter completed by a student in this trial resulted in a communication skills score provided by a standardized observer and a clinical skills score provided by a faculty physician. The total score was a simple addition of these two subscores. Clinical skills had a maximum of 30 points; communication skills had a maximum of 15 points. The combined scores were used as a measure of clinical performance. Combining clinical skills scores (history taking and physical exam) with communication skills scores is widely accepted and has been used by educators in the health professions for nearly two decades.¹⁶

Each student's mean communication score awarded by the standardized observers was correlated with the mean clinical skills scores awarded by the faculty. Descriptive statistics

and univariate analysis using NCSS[®] software (NCSS, Kaysville, UT) summarized the communication and clinical skills scores and student feedback related to their interactions with faculty.

After completion of an observed encounter and receiving feedback from the standardized observer and faculty in the pediatrics clinic, participating students were asked to complete a written feedback questionnaire. This instrument included four questions about student perceptions related to their interactions with the faculty.

The mean SOCE scores (after arcsine transformation to normalize the data¹⁷) of the participating students were correlated with the mean scores the students obtained in an OSCE completed during the pediatric clerkship using a Pearson correlation. The OSCE for the pediatric clerkship involved three cases. For each case, a student could spend up to 15 minutes with the standardized patient obtaining a medical history, performing a physical exam, and providing information. Two of the cases were built around clinical scenarios involving young children (eg, fever, vomiting, wheezing) and the third case involved an adolescent with a clinical concern (eg, headache, knee pain). The young-child cases employed adult standardized patients who played the role of the child's parent; the adolescent case had an adolescent standardized patient. The OSCE cases used a clinical skills checklist and the same communication skills checklist used by the standardized observers.

Students completed the OSCE midway through their pediatric clerkship and within two weeks of their standardized observer sessions. Consequently, at the time of a student's participation in the SOCE, approximately half of the students had already completed the OSCE and the other half had not.

Generalizability study

The reliability of the clinical skills scores and the overall scores (a combined clinical skills and communication skills scores) were calculated using generalizability theory and the G_String II[®] implementation of urGENOVA[®] software. (Robert Brennan, Center for Advanced Studies in Measurement and Assessment, Iowa City, IA) The variance components were calculated with the clinical encounter facet nested within students, or an encounter-nested-within-student [encounter: student] random model. A D-study was undertaken to predict the reliability of the performance scores based on different numbers of observed clinical encounters completed by a student.

Results

During the period of study, a total of 199 clinical encounters were observed by standardized observers, and faculty completed 197 of the 199 rating forms following the students' verbal presentations. A total of 51 third-year medical students in the General Pediatrics Clinic participated in this project. The mean number of SOCEs per student was 3.9, with a range of 2–12; 41 students completed three or more SOCEs. During half-day clinic sessions, between one and three clinical SOCEs were observed per student.

A total of 15 faculty participated in the clinical skills ratings and completed a mean of 13.3 rating forms (range 1–22 per faculty and a median 12). The range of scores from the six items on the rating scale was 19–30 out of a maximum total of 30. The mean clinical skills score was 26.5 (standard deviation [SD] 2.6), and the median score was 27. The range of scores on the 15-item communication checklist was 11–15; the average score was 14.5 (SD of 0.7), and the median score was 15. The mean of the combined faculty and standardized observer scores was 41.0 (SD of 2.7) The combined scores for the students ranged from 32 to 45 out of a maximum total of 45 points with a median of 41. The correlation between each students average clinical skills scores awarded by faculty and communication scores standardized observers student was small ($r = 0.28$) and not statistically significant ($P = 0.13$).

Reliability and validity

Approximately 30% of the score variance in the total SOCE scores (combined standardized observer and faculty ratings)

was attributable to the student facet in the generalizability analysis (Table 1). Additionally, the calculated reliability of the total SOCE scores (derived from summing the clinical skills rating scale and communication checklist) based on 10 case observations per student is 0.81 (Table 2). There was a moderate positive correlation between the average scores obtained by students on their SOCEs and their OSCE cases ($r = 0.54$; $P < 0.01$) completed during the same pediatrics clerkship.

Student perceptions

We successfully collected information from 50 of the 51 students following 191 of the 199 observed student–patient encounters. In describing student feedback, all percentages are calculated using 191 as denominator. Following 188 (98.4%) of the SOCEs, students agreed or strongly agreed that the faculty paid attention to their presentation about the patient. Following 165 (86.4%) of the SOCEs, students agreed or strongly agreed that the faculty person provided them with feedback on the history they collected from the patient. Following 159 (83.2%) of the SOCEs, students agreed or strongly agreed that their supervising faculty had provided feedback on their clinical reasoning skills. Following 185 (6.8%) of the SOCEs, students agreed and strongly agreed that the feedback they received on their clinical skills was helpful for improving these skills.

Discussion

Our data suggest that the SOCE produces reliable scores for assessing medical students' performance during real clinical encounters with pediatric patients. This should not be surprising. van der Vleuten and colleagues have demonstrated that the primary driver of reliability in performance assessment is adequate sampling of student performance and not standardization of the cases used for the assessments.¹⁸ Researchers have also concluded that the mini-CEX and the Long Case, both of which utilize clinical encounters, are reliable assessment methodologies.^{13,19} Our work suggests that the SOCE has similar reliability.

Table 1 Generalizability analysis using urGENOVA[®] for estimating the sources of skills score variance

Facets	Degrees of freedom	Variance component	Percentage of variance
Student	51	2.1886	30.5%
Encounter:student	146	4.9985	69.5%

Notes: Data were collected by observing 51 third-year medical students on the pediatrics clerkship at the University of Iowa Carver College of Medicine during the 2006–2007 academic year

Table 2 Predicted systematically observed clinical encounter (SOCE) score reliability as a function of the number of student-patient encounters completed by students

Number of SOCE cases	Score reliability
2	0.467
3	0.568
4	0.637
5	0.686
6	0.724
7	0.754
8	0.778
9	0.798
10	0.814
11	0.828

Notes: The generalizability study was undertaken with urGenova[®] using all available data collected by observing 51 third-year medical students on the pediatrics clerkship at the University of Iowa Carver College of Medicine during the 2006–2007 academic year

Further development of performance-based assessment methodologies is important for medical educators. The OSCE does not fully meet the assessment needs of a pediatric clerkship at our institution because the use of infants and young children as standardized patients violates child labor laws in Iowa.²⁰ Although this legal restriction does not apply to all medical schools, more universal ethical concerns have been raised about using children in OSCEs.^{21,22} In addition, it is difficult to standardize the performances of infants and children.²²

Although the mini-CEX and the Long Case can be used to assess the skills of medical learners with infants and children and are psychometrically sound, both of these methods require considerable faculty time for undertaking observations of students. Currently, physician faculty infrequently observe medical learners examining patients at most United States academic centers due to the lack of time and competing clinical demands on faculty.²³ In an earlier report we showed that lay standardized observers can assess the communication skills of medical students while engaged in patient care.¹⁵ In the current report we show that combining the communication skills ratings of standardized observers with physician faculty ratings of the learners' data collection and clinical reasoning skills results in a reliable measure of clinical performance.

The SOCE, like the mini-CEX and the Long Case, does not rely on a checklist for assessing clinical skills but relies on physician judgment. This is in contrast with the OSCE where predetermined checklists of questions and physical exam maneuvers are typically used. The absence of checklists for evaluating the clinical skills of learners should not be viewed as a weakness. Considerable research suggests

that using physician ratings in assessing clinical skills has advantages. Physician ratings are known to be at least as reliable as checklists²⁴ and may be more sensitive to different levels of expertise.²⁵ Physician ratings also provide insight into a more broadly based set of skills compared with checklists which are highly content specific.²⁶

We also think it important to consider the positive impact the SOCE can have on learners by facilitating formative feedback. In contrast to the OSCE, the SOCE has no checklist to keep hidden from the learner, which allows faculty to be very specific in their feedback. The students participating in our project reported that they almost always obtained feedback on their clinical skills from the faculty when participating in the SOCEs. Previously we reported that students routinely received helpful formative feedback about their communication skills from the standardized observers participating in this project.¹⁵ Therefore, the SOCE is not only reliable and feasible, it supports learning.

Our research has several limitations. The first is the limited number of encounters which we observed. A larger sample size would provide a more accurate estimate of the reliability of the scores arising from this assessment method. However, our estimates of the reliability of the SOCE are consistent with reliability estimates for other performance assessment methods using real clinical encounters. Another limitation of the study is that our observations were all completed in a single pediatrics clinic, and we do not know how our findings generalize to other clinics. There also remain questions about the validity of the scores generated by the SOCE. While we report a significant correlation between students' SOCE and OSCE scores, further study is needed.

Conclusion

At this time it appears the SOCE may join the mini-CEX and the Long Case as reliable assessment methodologies that can be used in real clinical settings. Our data suggest that students' clinical performances can be reliably assessed by completing 10 SOCEs. Whether or not this approach allows expansion of performance-based assessment in clinical settings deserves further study.

Acknowledgment

The authors thank Jo Bowers for her indispensable assistance in coordinating the standardized observers on this project.

Funding

This project was supported by the Educational Development Fund, Office of Consultation and Research in Medical

Education, UI Carver College of Medicine and the William and Sondra Meyers Family Professorship, UI Carver College of Medicine.

Disclaimers

Presented in part at the 13th Ottawa International Conference on Clinical Competence, Melbourne, Australia, March 2008.

Ethical approval for this research project was obtained from the University of Iowa Institutional Review Board.

Disclosure

The authors report no conflicts of interest in this work.

References

- Newble DI, Norman G, van der Vleuten C. Assessing clinical reasoning. In: Higgs J, Jones M, editors. *Clinical Reasoning in the Health Professions*. 2nd ed. Oxford, UK: Butterworth-Heinemann. 2000: 168–178.
- Holmboe ES. Faculty and the observation of trainees' clinical skills: problems and opportunities. *Acad Med*. 2004;79:16–22.
- Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ*. 1979;13:41–54.
- Hodges B. Validity and the OSCE. *Med Teach*. 2003;25:250–254.
- van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ*. 2005;39:309–317.
- Hodges B. OSCE. Variations on a theme by Harden. *Med Educ*. 2003;37:1134–1140.
- Bligh J, Bleakley A. Distributing menus to hungry learners: can learning by simulation become simulation of learning? *Med Teach*. 2006;28:606–613.
- Teoh NC, Bowden FJ. The case for resurrecting the long case. *BMJ*. 2008;336(7655):1250.
- Norman G. The long case versus objective structured clinical examinations. *BMJ*. 2002;324(7340):748–749.
- Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA*. 2009;302:1316–1326.
- 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 Suppl 10:S56–S60.
- Kogan JR, Bellini LM, Shea JA. Implementation of the mini-CEX to evaluate medical students' clinical skills. *Acad Med*. 2002;77: 1156–1157.
- Wass V, Jones R, van der Vleuten C. Standardized or real patients to test clinical competence? The long case revisited. *Med Educ*. 2001;35: 321–325.
- Wilkinson TJ, Campbell PJ, Judd SJ. Reliability of the long case. *Med Educ*. 2008;42:887–889.
- Bergus GR, Woodhead J, Kreiter CD. Trained lay observers can reliably assess medical students' communication skills. *Med Educ*. 2009; 43:688–694.
- Swanson DB, Norman GR, Linn RL. Performance-based assessment: lessons from the health professions. *Educ Res*. 1995;24:5–11, 35.
- Kleinbaum DG, Kupper LL, Muller A, Nizam KE. *Applied Regression Analysis and Other Multivariable Methods*. 3rd ed. Pacific Grove, CA: Duxbury Press; 1998:252.
- van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ*. 2005;39:309–317.
- 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.
- Under fourteen – permitted occupations, Child Labor. Code of Iowa. Vol § 92.3; 2003.
- Hilliard RI, Tallett SE. The use of an objective structured clinical examination with postgraduate residents in pediatrics. *Arch Pediatr Adolesc Med*. 1998;152(1):74–78.
- Tsai TC. Using children as standardised patients for assessing clinical competence in paediatrics. *Arch Dis Child*. 2004;89:1117–1120.
- Howley LD, Wilson WG. Direct observation of students during clerkship rotations: a multiyear descriptive study. *Acad Med*. 2004;79: 276–280.
- 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*. 1998; 73:993–997.
- Hodges B, McLroy JH. Analytic global OSCE ratings are sensitive to level of training. *Med Educ*. 2003;37:1012–1016.
- Regehr G, Freeman R, Hodges B, Russell L. Assessing the generalizability of OSCE measures across content domains. *Acad Med*. 1999; 74:1320–1322.

Appendix

Appendix I Clinical skills rating scale completed by faculty physicians after each SOCE. This instrument has six rating scales about the student's data collection and clinical reasoning skills. A student's score can range between 6 and 30 points

Domain scores					
History	1	2	3	4	5
Completeness of HPI: Did the student present all of the major facts of the HPI?	Few	Some	Many	Most	All
Development of HPI: Did the student develop the HPI by focus on onset of symptoms, site, time course, severity, setting, and aggravating-relieving factors?	Little	Limited	Partial	Consistent	Complete
Other appropriate history: Did the student provide other history (past medical history, family history, social history, ROS) appropriate to the presenting problem?	Little	Some	Many	Most	All
Accuracy of history: Did the student present any incorrect information?	Large amount	Some	Little	Rare	None
Physical exam	1	2	3	4	5
Completeness of PE: Did the presentation include appropriate detail about the important physical findings?	Little	Some	Many	Most	All
Clinical reasoning	1	2	3	4	5
Soundness of reasoning: Development of Diff DX and problem list demonstrated sound clinical reasoning.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

Abbreviations: Diff DX, differential diagnosis; HPI, history of the present illness; PE, physical exam; ROS, review of systems; SOCE, systematically observed clinical encounter.

Advances in Medical Education and Practice

Dovepress

Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied healthcare professional education. The journal covers undergraduate education, postgraduate training and continuing medical education

including emerging trends and innovative models linking education, research, and healthcare services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>