Peer reviewed article

Evaluation of unstructured medical school examinations: prospective observational study

Andreas Zeller, Manuel Battegay, Niklaus Gyr, Edouard Battegay

Medical Outpatient Department, University Hospital, Basel, Switzerland

Summary

Aim: To evaluate the final examination in Ambulatory General Internal Medicine of the Medical School of the University of Basel, Switzerland regarding students' performance rated by examiners and patients, examiners' and students' self-assessment and examiners' performance concerning fairness and difficulty of the examination rated by students and patients.

Method: Prospective observational study of 144 Medical students judged by 29 pairs of examiners. Students examined 66 real untrained outpatients. Assessment by questionnaire during an unstructured final Medical School examination. Marks could be given between 1 (= very poor) to 6 (= very good). Fairness and difficulty of the examination was measured by visual analogue scale (1 to 10).

Results: Patients judged students' performance better than examiners (5.45 ± 0.46 vs. 5.22 ± 0.65, p = 0.005). Examiners assessed students performance better than the students themselves (5.22 ± 0.61 vs. 4.91 ± 0.54, p = 0.001). Patients considered examiners as having examined fairly in 84.6%, and students rated examiners as having examined fairly on visual analogue scale (1.29 ± 1.75). Students and examiners judged the exam to be similarly difficult (5.97 ± 1.76 vs. 5.92 ± 1.14, p = 0.77, r = 0.72).

Conclusion: An unstructured Medical examination – the long case – provides consistent results which are accepted as fair by the students, patients and examiners. Patients and examiners judge students' performance more benignly than students themselves. Examiners and examinees consider the long case as serving a meaningful purpose regarding assessment of clinical competence and doctor/patient relationship .

Key words: OSCE; medical education; medical examination

Introduction

Long cases, ie, assessing real and untrained patients in an unstructured manner, are being replaced by structured examinations such as the objective structured clinical examination (OSCE) [1] in undergraduate clinical examinations [2]. The OSCE has been found to be a valid and reliable method for assessing clinical knowledge [3]. OSCE performance also showed a significant correlation between interpersonal skills and clinical competence [4]. These findings elicited a hot and controversial debate about the appropriateness of the long case in graduate and licensing examinations [5–7].

The final examination of the Medical School of the University of Basel in General Internal Medicine, particularly in ambulatory care, is an unstructured clinical examination of real and untrained patients recruited from the Medical Outpatient Department, ie, a long case. We aimed at evaluating this examination regarding (1) students' performance rated by examiners and patients; (2) examiners' and examination's performance rated by students and patients; (3) examiners' and students' self-assessment; (4) Patients', students', and examiners' general assessment of the examination.

Participants, methods, and results

No financial support declared.

Between 130 and 150 Medical students take their final medical exam each year at the Medical School of the University of Basel, Switzerland. The final exam consists of more than a dozen parts covering the whole spectrum of Clinical Medicine in different formats including multiple choice exams and unstructured oral examinations. One of these exams, the one in Ambulatory General Internal Medicine, is given at the Medical Outpatient Department and has been investigated in this study. This particular part of the final examination aims at testing the doctor patient relationship, data gathering, communication, diagnostic and management skills. The exam is divided into three parts; an initial part with 10 minutes of observed history-taking, 60 minutes of unobserved data gathering and clinical examination, and a final 20 minutes of presentation and discussion of the patient's problems and potential management.

We prospectively investigated all 29 teams of examiners (each team consisting of a member of the Medical Faculty and a General Practitioner) and 144 students taking their final examination at the Medical School of the University of Basel, Switzerland from October until November 2001. Sixty-six patients of our Medical Outpatient Department were recruited and agreed to participate as untrained exam patients.

Examiners, students, and patients were asked to complete a questionnaire immediately after the examination with 5 items for the examiner (assessing students' performance, self estimation of fairness, judging difficulty of the exam), 5 items for students (self assessing of their own performance, judging the difficulty of the exam, judging fairness of examiners) and 3 for patients (judging students' performance, difficulty and fairness of the exam respectively). Additionally examiners, students and patients were enquired to judge the course of the examination in its existing form. Marks could be given between 6 (= very good) and 1 (= very poor) corresponding to standard Swiss practice in schools and exams (4 = being sufficient, ie, adequate for passing the exam). Questions concerning appraisal of the exam and examiners were assessed by a visual analogue scale (ranking from 1 to 10 cm, see figure 1). Examiners and students were asked to self-assess their own presentation before and just after the exam. Patients were asked to choose the most appropriate description of the examiner (too severe, severe, fair, unfair, mild, too mild). The failure rate is extremely low in this final exam after 6 years of Medical School on the average 1% of students failing the entire exam.

Numbers given were means ± standard deviation (SD). Qualitative parameters were given as proportion (percentage). Differences were calculated by student's t-test and correlation by Pearson Product Moment Correlation. Statistical tests were performed with GB-STAT® for Windows, Version V6.0, Dynamic Microsystem Inc. (Silver Spring, MD, USA).

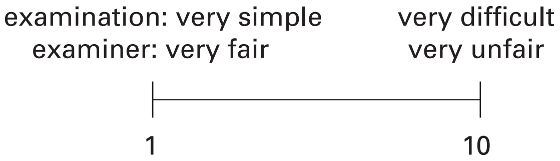


Figure 1

Use of visual analogue scale to assess severity of examination and fairness of examiner.

Results

A total of 144 students (76 male, 68 female) took the final exam. Patients judged students to perform significantly better in the exam than examiners (patients vs. Faculty members: 5.45 ± 0.46 vs 5.22 ± 0.65, respectively, p = 0.005; patients vs. General Practitioners: 5.45 ± 0.46 vs. 5.22 ± 0.58, p = 0.003).

Self assessment of students was not altered significantly during the exam but marks tended to decrease after the exam (marks before the exam: 4.98 ± 0.46; marks after the exam, but before knowing the mark given by the examiners: 4.91 ± 0.54, p = 0.051, r = 0.59). Examiners issued significantly better marks than students self assessed their performance in the examination, but there was not a high degree of individual correlation between the students self-assessment and the assessment of the examiners (5.22 ± 0.61 vs. 4.91 ± 0.54, p = 0.001, r = 0.46). There was a high correlation between appraisal by members of the Faculty and general practitioners (5.22 ± 0.65 vs. 5.22 ± 0.58, r = 0.83).

Patients considered examiners to examine fairly in 84.6% of all exams, severely in 9.1%, and mildly in 6.3%. None of the patients considered examiners to examine unfairly, too severely or too mildly. On visual analogue scale (1 = very fair and 10 = very unfair) students rated examiners as examining very fairly (1.29 ± 1.75), while examiners

Table 1

Marking of students, estimation of examination's severity and rating of fairness of examiners.

	examiners (n = 58)	patients (n = 66)	students (n = 144)	p value
marking (mean ± SD)*	5.22 ± 0.62	5.45 ± 0.46	4.91 ± 0.54[#]	p <0.05
severity of examination (visual analogue scale)**	5.92 ± 1.13	–	5.97 ± 1.76	0.77
fairness of examiner (visual analogue scale)[+]	2.6 ± 1.80[#]	–	1.3 ± 1.70	p <0.05

* ranking 1 to 6, 6 = excellent, 5 = good, 4 = sufficient, 1 = very poor

** visual analogue scale1 to 10, 1 = very simple, 10 = very severe

[+] visual analogue scale 1 to 10, 1 = very fair, 10 = very unfair

[#] self-estimation by student and examiner respectively assessed themselves as having examined significantly less fairly (2.56 ± 1.30, p = 0.001). On visual analogue scale (1 = very simple and 10 = very difficult) students judged the exam to be of similar difficulty as the examiners themselves (5.97 ± 1.76 vs. 5.92 ± 1.14, p = 0.77, r = 0.72).

Patients described students behaviour as calm (42.3%), confident (26.8%), competent (18.3%), nervous (9.2%) and stressed (3.4%). Nearly half of students (44.4%) felt they were "normally" nervous, 37.3% stressed, 11.3% extremely nervous and 7% calm just before the exam.

About two thirds of students and examiners (69.7%, respectively 63.8%) felt that the exam in its existing form serves a meaningful purpose concerning testing clinical competence and doctor/patient relationship, one fifth of students (21.1%) and one third (34.5%) of examiners judge the exam as a suitable form to assess students clinical competence. One examiner (1.7%) and 13 students (9.2%) thought that the exam does not serve a meaningful purpose.

Comment

Our study investigated an unstructured examination with a real patient, the so-called long case. The debate on the reliability of an unstructured oral examination has raised legitimate questions and controversies [5–8]. Such examinations do not seem to achieve reasonable levels of reproducibility [9]. The objective structured clinical examination (OSCE) has improved the reliability of oral exams [1, 10]. However, the OSCE only examines various "fractured" components of a real assessment of patients. Use of standardised or simulated patients (actors) to improve reproducibility may limit the complexity of real medical situations that can be depicted.

This study has investigated an examination with 144 examinees, 58 examiners and 66 untrained real patients. Patients judged the students' performance better that examiners had assessed it. Possibly, patients unconsciously identify with the students under stress of the examination and therefore gave better marks. Patients may feel supportive, also of poorly-performing examinees, and therefore may judge aspects of the patient-doctor relationship in a different way from examiners. Students estimated that their performance was significantly worse than examiners and patients. A reason for this may be low self esteem, conscious understatement or false expectation of the required knowledge. Interestingly, the exam did not change students' self-assessment before and just after the exam.

Further, our study suggests that this long case examination was judged to have a high degree of fairness. This finding is based on independent assessment by patients and students. A possible limitation for the students' assessment concerning fairness is that none of the students failed the specific examination investigated in our study. This may have contributed to the students' positive assessment of the examination, inclusively fairness. To obviate this point, students were asked to complete the questionnaire after the exam but before they were given the final mark. Examiners judged the administration of the exam more harshly and estimated themselves to be less fair. Because examiners are more likely to determine the Medical School's policy concerning examinations, this critical judgement of the "long case" may lead to a complete suppression of the long case examination. The high degree of concordance concerning the difficulty of the exam may allow the conclusion that examiners can correctly assess their own performance concerning the level of difficulty of the exam, ie, examiners can correctly evaluate where the students stand. As noted elsewhere [7] some examiners remarked that the long case offers a short but worthwhile opportunity to the examiner to give the student last tips and advice for his future function as a young doctor.

The long case, in contrast to the OSCE, creates a situation of daily life in Medicine with real untrained patients and their presenting complaints in an individual real world manner. Asking examiners and examinees about the meaningful purpose of this type of examination we wanted to know whether the exam in its existing form (long case) is a suitable form to test clinical competence and doctor/patient relationship. Most students and examiners felt that the exam served a meaningful purpose. Nevertheless one fifth of students and one third of examiners thought that the exam only partially serves a meaningful purpose. In particular, examiners mentioned the lack of standardisation of the exam. Therefore, according to Norcini [5] we believe that the long case without modifications such as direct and uninterrupted observation of student patient interaction should probably not be used to make critical decisions about the competence of a student. Our setting includes 10 minutes of observation of the student's history-taking by the examiners and this has an important impact on final marking. The observation of uninterrupted history taking is a useful and valid supplementary tool to measure a student's clinical competence [11]. Obviously, guidelines should attempt to structure unstructured examinations as far as possible in order to eliminate inequalities of examinations. But, this should not lead to the complete elimination of the long case, ie, the real patient form of Medical examination.

Our study has some limitations. The teams of examiners filled out the questionnaire in the same room, therefore we cannot perform a proper inter-examiner comparison. The patient recruitment may have selected highly motivated patients who participated voluntarily and did not receive any reimbursement. Hence, this population of patients may have regarded students and examiners more favourably than an "unbiased common patient".

In summary, the long case, particularly in final examinations, represents a tool to assess clinical competence of medical students, which is accepted as being fair by students, examiners, and patients. Patients and examiners judge students' performance more benignly than students themselves and most agree that the exam serves a meaningful purpose regarding assessment of clinical competence and doctor/patient relationship.

Correspondence
Andreas Zeller, MD
Medical Outpatient Department
University Hospital Basel
CH-4031 Basel
E-Mail: zellera@uhbs.ch

References

1 Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). Med Educ 1979;13:41–54.

2 Hardy KJ, Demos LL, McNeil JJ. Undergraduate surgical examinations: an appraisal of the clinical orals. Med Educ 1998; 32:582–9.

3 Sloan DA, Donnelly MB, Schwartz RW, Strodel WE. The Objective Structured Clinical Examination. The new gold standard for evaluating postgraduate clinical performance. Ann Surg 1995;222:735–42.

4 Colliver JA, Swartz MH, Robbs RS, Cohen DS. Relationship between clinical competence and interpersonal and communication skills in standardized-patient assessment. Acad Med 1999;74:271–4.

5 Norcini JJ. The death of the long case? BMJ 2002;324:408–9.

6 Wass V, Jones R, Van D, V. Standardized or real patients to test clinical competence? The long case revisited. Med Educ 2001; 35:321–5.

7 Meadow R. The structured exam has taken over. BMJ 1998; 1329.

8 Van D, V. Validity of final examinations in undergraduate medical training. BMJ 2000;321:1217–9.

9 Norcini JJ. The validity of the long case. Med Educ 2001; 35:735–6.

10 Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical examination. Med Educ 1988; 22:325–334.

11 Wass V, Jolly B. Does observation add to the validity of the long case? Med Educ 2001;35:729–34.

Swiss
Medical Weekly

Official journal of
the Swiss Society of Infectious disease
the Swiss Society of Internal Medicine
the Swiss Respiratory Society

The many reasons why you should choose SMW to publish your research

What Swiss Medical Weekly has to offer:

- SMW's impact factor has been steadily rising, to the current 1.537
- Open access to the publication via the Internet, therefore wide audience and impact
- Rapid listing in Medline
- LinkOut-button from PubMed with link to the full text website http://www.smw.ch (direct link from each SMW record in PubMed)

- No-nonsense submission – you submit a single copy of your manuscript by e-mail attachment
- Peer review based on a broad spectrum of international academic referees
- Assistance of our professional statistician for every article with statistical analyses

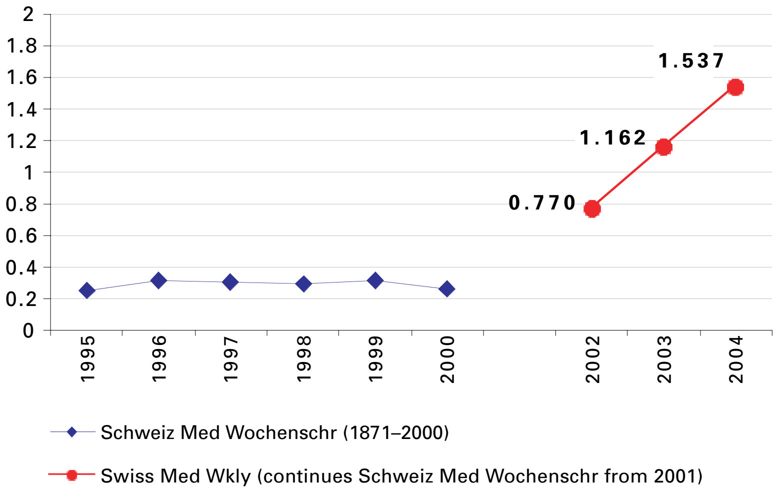
- Fast peer review, by e-mail exchange with the referees
- Prompt decisions based on weekly conferences of the Editorial Board
- Prompt notification on the status of your manuscript by e-mail
- Professional English copy editing
- No page charges and attractive colour offprints at no extra cost

We evaluate manuscripts of broad clinical interest from all specialities, including experimental medicine and clinical investigation.

We look forward to receiving your paper!

Guidelines for authors:
http://www.smw.ch/set_authors.html

Impact factor Swiss Medical Weekly



EMH FMH SCHWABE
Editores Medicorum Helveticorum

All manuscripts should be sent in electronic form, to:

EMH Swiss Medical Publishers Ltd.
SMW Editorial Secretariat
Farnsburgerstrasse 8
CH-4132 Muttenz

Manuscripts:	submission@smw.ch
Letters to the editor:	letters@smw.ch
Editorial Board:	red@smw.ch
Internet:	http://www.smw.ch