Peer reviewed article

Assessing clinical competence

A pilot project to evaluate the feasibility of a standardised patient-based practical examination as a component of the Swiss certification process

Nu V. Vu[a], *Anne Baroffio*[a], *Philippe Huber*[a, b], *Carine Layat*[a], *Margaret Gerbase*[a, c], *Mathieu Nendaz*[a, c]

[a] University of Geneva Faculty of Medicine, Unit of development and research in medical education, Geneva, Switzerland
[b] University of Geneva Faculty of Medicine, Department of Rehabilitation and Geriatrics, Geneva, Switzerland
[c] University of Geneva Faculty of Medicine, Department of Internal Medicine, Geneva, Switzerland

Summary

Purpose: advances in the assessment of clinical competence have prompted medical schools and licensing authorities to complement written and oral tests with practical ones. The purposes of this project were to (1) determine how clinical competencies not effectively addressed on the present Swiss federal examinations can be assessed adequately on a standardised patient-based practical examination (SCE) and (2) evaluate the SCE validity, reliability and feasibility.

Method: a bilingual, three-hour standardised patient-based clinical examination was pilot tested in 2003 with 48 volunteered fifth and sixth-year students from the five Swiss medical schools. All students took the same eight 15-minute patient cases. To ensure the test content validity, test cases were selected by a multi-disciplinary and -institutional committee of clinical faculty on the basis of predefined exam blueprint criteria and in reference to the Swiss catalogue of learning objectives.

Results: moderate correlations between the SCE and the existing Federal final written examinations (0.46) and the newly pilot-tested structured oral examination or SOE (0.56) [3] suggested that they were complementary to one another and that each might emphasise aspects of the clinical competence which others might not. The reliability (a coefficient) of the SCE scores ranged from 0.73 to 0.77.

Conclusions: limited experiences gathered throughout the SCE pilot project demonstrated its feasibility. Preliminary results suggested that SCE scores had a good level of construct validity and reliability and seemed to complement scores obtained on the final certification written examinations and the newly tested SOE. These results, however, need to be further confirmed with larger samples studies.

Key words: assessment; clinical competence; performance-based examination; standardized patients; certification

Background

With the recent adoption of the Swiss Catalogue of Learning Objectives for Undergraduate Medical Training (2002) [1] and in anticipation of the upcoming introduction of the New Federal Regulation of the Undergraduate Training in the Medical Professions (LPMed) [2], the Joint Commission of the Swiss Medical Schools (CIMS) has initiated a round of discussions on the future standing of the federal certification process. At present, the final federal examinations consist of a three-step approach to certification. The first includes five written examinations taken at the end of the 3rd year of the six-year medical curriculum. The second step is a series of 10 oral and 5 written multiple-choice specialty examinations taken either at the 5th or 6th year of medical school. The third step, taken at the end of the 6th year, consists of three oral long case examinations, each in Internal Medicine, Surgery and Paediatrics.

With the introduction of the LPMed, it is projected that the certification process will limit itself to one final examination, comprehensive in nature and administered at the end of the medical program. With this consideration, the procedure leading to the final certification examination, hence the students' progression through the medical school will rely mainly on each faculty policy and requirements. Therefore, the new federal examination represents a final and comprehensive "verification" process and as such, it should address the candidates' clinical competence at the end of their medical school as well as their readiness to start

This study has been supported by a grant from the Helmut Horten Foundation and the Federal Office of Public Health, and with the support from the Faculty of Medicine of Basle, Bern, Geneva, Lausanne & Zurich. The opinions expressed here are those of the authors and should not be attributed to the funding agencies and institutions.

their residency training. Towards this purpose, it is proposed that the multiple-choice examination format should be maintained for the final examination with a few needed revisions and improvements. On the other hand, skills which are at present not directly or adequately assessed with the multiple-choice formats should be complemented with other examination formats. Two were proposed for consideration and to be pilot tested for feasibility within the context of the Swiss certification process. One is the structured oral examination (SOE), a more standardised format of the oral exam which was proposed to replace the present one. The other is the standardised patient-based clinical examination (SCE) proposed to complement the existing written and oral examinations in assessing competencies not directly or well evaluated by these two formats. Overall, the future final certification examination will consist of the written exam which could be complemented with a structured oral and/or a standardised clinical exam. The design, development and pilot-testing of the SCE in the context of the Swiss final certifying examination are the focus of the present paper. The design, development and pilot-testing of the SOE are presented and detailed in a separate paper [3].

Most developments in the area of clinical skills and clinical competence evaluation have taken place since the 60's. However, only in the last 15 years in several medical schools in North America have senior medical students been required to pass a clinical practical examination in order to graduate [4–7]. During that same period, the use of standardised patient-based clinical examinations for the purposes of certification and licensing has also been pilot tested and introduced. This type of examination, also referred at times as the Objective structured clinical examination or OSCE, was first described by Harden et al. [8]. As used nowadays, it consists of a series of stations represented by clinical cases simulated by a standardised patient or SP, in which candidates are asked to perform a series of clinical tasks while being recorded and/or scored on checklists and rating scales respectively by trained SPs and/or physicians. A simulated or standardised patient is a lay person carefully coached by a specific training method to simulate accurately and in a standardised manner an actual patient [9]. Of note are four existing large-scale certification and licensing initiatives carried out in Canada and the United States.

The first licensing initiative was carried out in Quebec, Canada. Starting in 1990, all family physicians who would want to setup their practice in the province of Quebec would need to pass a standardised-patient based clinical exam in addition to the College of Family Physicians of Canada written and oral examinations [10, 11]. Analyses of the Quebec OSCE examination showed positive and good correlation coefficients between the OSCE and the short-answer management problems (0.56) and the simulated office orals (0.46). No score differences were found between candidates who took their exam in the English and French centres. Follow-up validity studies [12] indicated that high scores on the examination were significant predictors of competencies in consulting and prescribing, and mammography screening rates in initial primary care practice. The exam scores showed a sustained relationship over 4 to 7 years with indices of preventive care, and acute and chronic disease management [13].

This endeavour was followed in 1991 by the Medical Council of Canada (MCC) who completed its multiple-choice and key-feature short-answer licensing examination with a standardised patient-based clinical examination [14] referred to as the MCC Qualifying Examination Part II. Overall, the reliability of the overall exam varied between 0.61 and 0.78. The validity of the MCC performance examination has been demonstrated through the validity of its scoring, standard setting, and sequential testing approaches [15, 16].

The third undertaking was from the Educational Commission for Foreign Medical Graduates (ECFMG) which introduced in 1988 a standardised patient-based Clinical Skills Assessment (CSA) as a new requirement for foreign medical schools graduates seeking certification for entry into an accredited residency training program in the United States. Overall, validity studies of the CSA demonstrated that (a) the CSA "assessed proficiencies distinct from those assessed by the written USMLE *(United States Medical Licensing Exam)* and therefore provided evidence justifying its inclusion to the medical licensure process" [17], (b) there is a convergent validity between CSA scores on communications skills and clinical ability and the ratings assessing similar constructs on the mini-Clinical Evaluation Exercise [18], (c) the standardised patients' ratings of the candidates' communication skills were found to be valid and reproducible [19, 20], and (d) the holistic scoring was found to be a valid approach to score the candidates' post encounter patient notes [21]. Various studies demonstrated that the CSA scores have demonstrated good reliability coefficients (i.e. averaging from 0.70 to 0.80) [19, 22–23] and that similar levels of reliability were also found across multi-site and multi-language administrations of the CSA [22].

In a similar initiative, the National Board of Medical Examiners (NBME) had developed a prototype standardised patient-based clinical performance examination which was pilot tested from 1995 to 1998 with various medical schools [24–26]. With the preliminary results demonstrating the reliability of the CSE (0.77 to 0.82), the Federation of State Medical Boards and the NBME decided to add the CSE to the step 2 USMLE written examination [27]. Starting June 2004, all students trained in the United States are required to pass the examination, a prerequisite for residency training and licensure.

The validity and reliability of the standardised patient-based clinical assessment format and its

added values to the written, computerised and oral examinations, has been the basis for its consideration as a needed complement to the future Swiss certifying process. The purposes of this paper are to: (1) describe the development and pilot-testing of the standardised clinical examination (SCE) as a viable complement to the present federal final written examinations and the newly designed structured oral examination (SOE), (2) report preliminary results on its validity, reliability, and feasibility, and (3) assess correlations between the SCE scores with those on the written examinations and the SOE. The present study is not designed and intended to compare candidates' performances on the SCE either by their year or place of training.

Method

Test design and development

Test prototype – Anticipating various constraints of practicality and feasibility of administering the SCE as part of the Swiss certification process, and on the basis of previous findings related to its optimal test length [28, 29] and patient encounter or station duration [29–30], the SCE was conceived as a 2-hour examination with eight 15-minute patient encounters per student. The exam used simulated-standardised patients (SPs) to portray the patients of its clinical cases. For each case, the candidates were given up to 15 minutes for an appropriate clinical encounter with the patient while carrying out specific clinical tasks required by the case. Depending on the objectives of each case, the candidates are assessed on various clinical skills. These include ability to perform a focused history-taking and physical examination, to provide a relevant feedback and counselling to the patient, to derive accurate diagnostic hypotheses and propose an appropriate follow-up diagnostic and/or management plan for the patient. In addition to these skills, the candidates were evaluated across all cases on their communication skills, interpersonal relationship, and overall performance. For practical purposes, no paediatric patients were included, and pelvic, rectal and female breast exams were not part of the physical examination. Given that the Swiss certification exam has always been administered in two languages, the pilot SCE was developed in a German and a French version.

Test blueprint – Given the intent that the SCE is introduced to complement the written multiple-choice final exam and the new pilot tested SOE, it specifically focused in assessing how senior medical students perform their clinical skills as well as how they communicate with and relate to the patient, skills which are presently not adequately assessed with the other two exam formats. The importance and necessity of assessing both of these skills were further confirmed by research findings suggesting a correlation between (a) the complementary cognitive (ie, clinical problem-solving) and non-cognitive (ie, communications) structures of clinical competence [5, 31] and (b) the relationship between non-cognitive competencies and quality of clinical care [12] and medical professionalism [32].

Given that the SCE was intended to be the final and formal verification of students' readiness to enter residency and hence to practice under supervision, the clinical cases and skills to be assessed on the SCE were derived from the Swiss catalogue of undergraduate learning objectives. To ensure the SCE test content validity, a blueprint was set up with predefined criteria based on which the clinical cases were selected. They included the following:

1) *Presenting complaint(s)* – Recommended for inclusion in the test are common presenting complaints with a clearly identifiable diagnosis, and a well-accepted, non-controversial diagnostic and management approach.

2) *Diagnosis* – Five main categories are identified from which the diagnoses are to be selected. They are: Cardiovascular / Respiratory; Digestive / Genitourinary; Neurological / Psychiatric; Constitutional symptoms (ie, hypertension, weight loss, obesity …); and Others (ie, ear, eyes, nose, throat, musculoskeletal, infections, immunology …)

3) *Primary and secondary clinical disciplines* involved in the cases. The disciplines were broadly categorised into Internal medicine, Surgery, Paediatrics, Obstetrics-Gynaecology, Psychiatry, and Family medicine.

4) *Type of care* which includes: acute, chronic, or follow-up

5) *Context of care* which includes: emergency, in-patient/hospital, outpatient/primary care

6) *Patient's age and gender*

Five of the eight cases were set up to assess the candidates' skills in history-taking, physical examination, elaborating accurate diagnostic hypotheses and proposing an appropriate follow-up diagnostic and/or management plan to the patient. Three cases were set up to assess the candidates' skills in history-taking, patient education or counselling in life style change or care giving, and proposing adequate follow-up diagnostic and/or management plan. All eight cases were set up to assess the candidates' oral communication and interpersonal skills.

Case selection, *authoring*, *translation and verification* – For the pilot project, a process for an inter-institutional and inter-disciplinary SCE case selection committee and its functioning process have been conceived and tested for its feasibility. Members of the SCE committee were selected from all the main disciplines to be covered on the exam blueprint and were identified from the Faculty of Medicine of Zurich, Berne and Geneva.

For reasons of practicality, efficacy and quality control of the test development process, each member of the case selection committee also served as first or second case author of one of the SCE cases and participated with the SP trainer in the training of the SPs of their respective case.

All cases exist in French and German versions. To ensure test validity and reduce variations among the translated versions, all translations were carried out by the same translator who has a medical background and whose native language is Swiss German and is fluent both in French and German. To ensure the accuracy of the translations, each case exam item translation was reviewed by its respective German or French speaking author or co-author.

Test administration

The 8-station exam was administered over one day for the 48 candidates and consisted of three 3-hour examining sessions with eight students scheduled per session. Overall, each candidate was scheduled for a three-hour examination during which they had eight 15-minute patient encounters. Before the start of the exam, an orientation

session was scheduled to guide the students through the examination.

Test candidates

While the SCE is intended for 6th year students, candidates for the pilot test were recruited among 5th and 6th year students because of the anticipated difficulties in recruiting them. Depending of the Faculty, students are in their elective year either in the 5th or 6th year and hence away from the campus. For the pilot test, 48 fifth- and sixth-year students-volunteers were recruited for the test site in Geneva (n = 24) and in Bern (n = 24). At each site, all candidates took both the SCE and the SOE [3]. For Geneva, the candidates were recruited from Lausanne (n = 5) and Geneva (n = 19), and for Bern the candidates were recruited from Zurich (n = 14), Bern (n = 9) and Basle (n = 1).

Test examiners

Given the difficulties in recruiting clinical examiners for a whole day pilot exam, the case author and co-author were asked to serve also as examiners of their SCE case. This process represented a limiting factor of the study given that in the real practice examiners are often not the original case author or co-author. For the present study, having the case author or co-author as examiners may further enhance the validity of the scoring process since they were familiar with the case objectives and the scoring instruments and process. For the exam in Bern, the examiners were from Bern and Zurich and for the exam in Geneva; all the examiners were from Geneva. The same set of examiners was used for all three of the exam sessions; in other words, each student had one examiner per case and for each case all students were examined by the same examiner. Before the start of the exam, a meeting was scheduled to orient the examiners to the examination and to review the scoring process.

Case and exam scores – All the SCE scores were calculated and reported in terms of percentage scores (score obtained/total score) and ranged from 0 (minimum) to 100 (maximum). For each case and the overall exam, five scores were calculated from the examiners' ratings. They include:

1) *CR* or *Clinical Reasoning score* – derived from checklist scores obtained on the following skills: history-taking, physical examination, patient education/counselling, diagnosis, and diagnostic investigation/management plan.

2) *COMM* or *Communications-Interpersonal Relationship score* – derived from a standard rating scale assessing the candidates' communication and interpersonal skills with the patient at the beginning, middle and end of the encounter,

3) *G* or *Global score* – The examiner was asked to rate on a scale from 1 to 5 how he/she appreciated the overall performance of the student.

4) *T* or *Unweighted total score* – It was the mean total of the CR and COMM scores.

5) T^w or *Weighted total score* – It was the mean total of the CR and COMM scores, with the CR scores having a weight of 2 and the COMM score a weight of 1. It should be noted that this differential weighting has been practiced in the last few years in the final Internal Medicine oral-practical exam in Geneva. This approach has been adopted on the basis that, in contrast to the COMM score which was based on one rating scale, the CR score was derived from several steps of the patient encounter and calculated from five checklist scores.

Statistical analyses

Descriptive statistics (mean, median and standard deviation) are derived for the competencies, case and exam scores. Pearson correlation was used to determine the relationships between the various test scores and Cronbach-alpha coefficient was calculated to derive the test score reliability or internal consistency.

Results and discussion

The SCE results were based on 46 fifth- and sixth-year medical students: 23 from the testing site in Geneva (Lausanne: n = 5; Geneva: n = 18) and 23 from the one in Bern (Zurich [n = 14]; Bern [n = 8]; Basle [n = 1]). One student from Geneva was late and did not complete the entire exam and hence was not included in the analysis. One student from Bern did not present to the examination.

SCE Case and CR components scores

Table 1 summarises the means and standard deviations of the case and overall exam CR scores and CR component scores which include the history-taking, physical examination, patient education/counselling, diagnosis, and diagnostic investigation/management plan scores. Overall, the CR component scores and hence the CR scores (42–87) assessed a wide range of performance and varied greatly in function of the patient cases. Furthermore, high or low performance on one CR component did not seem to result in similar performances on other components.

SCE CR, COMM, G, T and T^w scores

Table 2 summarises the means and standard deviations of the case and exam CR, COMM, G, T and T^w scores. Overall, the ranges of the case COMM scores (87–96) and to a certain extent the G scores (62–79) were more restricted than those of the CR (42–87), T (51–90) and T^w (54–81) scores. These results suggested that the SCE scores were case sensitive and was able to capture a wide range of the candidates' clinical performance. In addition, the scores seemed to be case specific and varied in function of the 8 patient encounters. As with previous research findings, these results implied that a valid and hence more accurate assessment of a candidate's clinical competency have to comprise a sufficient number of cases and observations of the candidate's performances. This implication is further substantiated by the relatively good reliability coefficients (Cronbach α) obtained for the CR (.63), COMM (.75), T (.68) and T^w scores (.76), and G ratings (.73).

Table 1

Means and standard deviations of candidates' (n = 46) SCE Clinical reasoning (CR) and CR component percentages scores on History, Physical examination, Patient education, Diagnosis, and Investigation & management.

Patient cases	CR Component scores					CR
	History-taking (n = 103)[a]	Physical Exam. (n = 38)[a]	Pat. Educ. (n = 21)[a]	Diagnosis (n = 16)[a]	Investigation and management plan (n = 19)[a]	
1. Diabetes Control	50 (15)	71 (15)	50 (19)	–	–	57 (14)
2. Persistent Cough	72 (12)	–	48 (17)	80 (29)	77 (33)	69 (17)
3. Jaundice in a newborn	55 (21)	–	70 (20)	64 (25)	61 (29)	62 (18)
4. Weight loss and insomnia	65 (14)	–	–	100 (0)	98 (8)	87 (6)
5. Acute abdominal pain	71(15)	83 (15)	–	59 (29)	62 (25)	68 (16)
6. Vaginal bleeding	43 (18)	58 (20)	–	36 (38)	32 (22)	42 (15)
7. Confusion	54 (16)	47 (19)	–	61 (33)	61 (31)	56 (14)
8. Dyspnoea on exertion	60 (18)	45 (14)	–	77 (25)	78 (31)	65 (13)
Overall CR component Score	**59 (19)**	**61 (23)**	**56 (22)**	**68 (35)**	**67 (33)**	**63 (14)**

[a] Category total number of checklist items on the exam

Table 2

Means, standard deviations, and reliability coefficient (α) of' SCE Clinical reasoning (CR), Communication (Comm), Global (G), Unweighted total (T) and Weighted total (T^w) score (n = 46).

	CR	COMM	G	T	T^w
1. Diabetes Control	57 (14)	92 (13)	73 (21)	66 (12)	66 (12)
2. Persistent Cough	69 (17)	89 (10)	79 (17)	73 (15)	68 (10)
3. Jaundice in a newborn	62 (18)	89 (13)	70 (17)	68 (15)	68 (15)
4. Weight loss and insomnia	87 (6)	96 (7)	66 (19)	90 (5)	81 (8)
5. Acute abdominal pain	68 (16)	90 (13)	74 (16)	72 (14)	75 (13)
6. Vaginal bleeding	42 (15)	87 (15)	62 (22)	51 (13)	54 (12)
7. Confusion	56 (14)	88 (13)	64 (18)	62 (12)	61 (9)
8. Dyspnoea on exertion	65 (13)	90 (15)	64 (17)	70 (11)	61 (11)
Overall Exam Score	**63 (8)**	**90 (8)**	**69 (11)**	**69(7)**	**67 (7)**
Alpha coefficient (reliability)	.63	.75	.73	.68	.76

Table 3

Pearson correlation coefficients between the CR component scores (n = 46).

	History-taking	Physical Exam.	Patient Education	Diagnosis	Investigation & management Plan
History-taking	1.00	0.25	0.09	0.27	0.22
Physical Exam.		1.00	0.45	0.00	0.00
Patient Education			1.00	0.22	0.40
Diagnosis				1.00	0.53
Investigation & Management				–	1.00

Inter-correlations between CR components scores (table 3) and between CR, COMM, G, T and T^w scores (table 4)

The above finding that high performance on one CR component of a case did not necessarily entail high performance on other components was further confirmed with the low to moderate inter-correlations between the CR components (.00 to .53). These results implied that the components might assess, to a certain extent, separate and independent competencies and therefore they are complementary in assessing the candidates' clinical reasoning. These correlations further substantiate findings concerning processes underlying the clinical reasoning [33] as well as our concept of representing the CR scores with the five component scores of history, physical examination, patient education/counselling, diagnosis, and diagnostic investigation/management plan scores.

Furthermore, moderate correlations between the CR and COMM scores (r = .53) implied that they are two different skills and hence validate our scheme to use and combine the CRS and COMM scores to derive the candidate overall T and T^w performance scores.

Correlations among the three overall exam performance scores revealed a relatively high correlation between the G and the T scores (.82) and a moderate one between G and the T^w scores (.66). In general, moderate size correlations were more of a typical finding in the literature [34–36]. This could be due to the fact that the total exam scores were often derived from checklist recordings and based on specific pre-defined criteria while the global ratings were based on examiners' overall impression and relied on criteria not defined or easily captured on the checklist. The results from this study seemed to support this explanation. One possible reason that the G score had a higher correlation with the T than the T^w score was that with

Table 4

Pearson correlations between the Standardised Clinical Exam CR, COMM, G, T and T^w scores (n = 46).

	CR	COMM	G	T	T^w
CR	1.00	0.53	0.65	0.78	0.94
COMM		1.00	0.48	0.47	0.62
G			1.00	0.82	0.66
T				1.00	0.77
T^w					1.00

the T^w, the objective scores are weighted two times higher than the subjective ones and hence may underestimate the subjective, non-checklist related aspects of the G score. The uses of checklist vs. global scores in standardised clinical performance examination have been much examined and debated. There is growing evidence on the value of adding global scores to the checklist performance evaluation. The rationale behind this proposition mostly relates to the complementarities of the two scoring methods. While checklist scores are highly content-specific, global scores may evaluate a more broadly based set of skills and approach to the patient [34–36]. Given this study's limited number of candidates and examiners, further replications are needed with bigger sample sizes before it can be determined which types of overall performance scores should be adopted for the final SCE.

Correlations between SCE, Structured Oral Exam (SOE) and Federal final multiple-choice (MC) written exam scores

The SCE was conceived and designed to complement the Federal final MC written exam and possibly the structured oral examination or SOE [3] by assessing skills which were not measured adequately by these two examinations. Scores from the SOE included the CR score representing the candidates' content knowledge and reasoning competency and the G score representing the examiners' ratings of the candidates' overall performance. A more detailed description of these two scores is provided in Hottinger et al. article [3]. For this study, the MC score was the candidates' average score on the final Part II five multiple-choice written examinations. The MC scores were provided confidentially by the Institut für Medizinische Lehre to the principle investigator for the purpose of research analyses.

Table 5 showed moderate Pearson correlations between the SCE T, T^w, and G scores and the SOE CR and G scores (R = 0.37 to 0.56), with the correlations between the SOE CR and the SCE T and T^w scores (R = 0.43 to 0.56) a little higher than the one between the two G ratings (R = 0.38). Similar range of correlations were found between the MC scores and the SCE T and T^w scores (0.53 and 0.46 respectively) and the G scores (R = 0.43). Because not all the SCE candidates had taken the written exam, the correlations were based on very limited number of subjects (n = 14) who were mainly from the Faculty of medicine in Geneva. Overall, these preliminary results suggested that while the three examinations total scores might measure similar competencies (i.e. content, reasoning) they may also be complementary and that each measured aspects of the competencies which others did not. Finally, the low correlation observed between the SCE COMM and the MC scores (r = 0.27) versus the moderate correlations found between the SCE CR scores and the MCQ (0.55) further contributed to the SCE construct validity by suggesting that the SCE did tap the candidates' communications competency which the MCQ by its design did not. However, due to the restricted number of subjects taking the written MC test, further replications are needed before this finding can be confirmed.

Exam costs

The cost of developing and administering a new 8-station bilingual exam was calculated to be around 1208 CHF per student or an average of 151 CHF per student and per "newly developed" station (vs. "existing" or already developed stations). This sum did not include however upfront costs related to the availability of a testing facility with equipped patient rooms, test administration

Table 5

Correlations between Standardised Clinical Examination (SCE) CR, COMM, T, T^w and G scores, the Structured Oral Examination (SOE) CR and G scores, and the Swiss Federal Part II Final multiple-choice (MC) written examination scores.

	SCE					SOE		MC
	CR (n = 46)	COMM (n = 46)	T (n = 46)	T^w (n = 46)	G (n = 46)	CR (n = 44)	G (n = 44)	MC (n = 14)
SCE CR	1.00	0.53	0.78	0.94	0.65	0.59	0.59	0.55
SCE COMM		1.00	0.47	0.62	0.48	0.55	0.51	0.27
SCE T			1.00	0.77	0.82	0.43	0.47	0.53
SCE T^w				1.00	0.66	0.56	0.53	0.46
SCE G					1.00	0.37	0.38	0.43
SOE CR						1.00	0.93	0.54
SOE Global							1.00	.48
MC								1.00

and analysis personnel, and conduct of research studies needed to upgrade and to control the exam quality and utility.

Conclusions

In summary, the experiences gathered throughout this pilot project demonstrated the feasibility of developing a bilingual standardised patient based clinical examination (SCE) to be complementary to the existing MC written tests, and administering it at two testing centres. The model introduced in this project to have a multi-institutional and multi-disciplinary exam committee, working on commonly defined exam blueprint criteria to select patient cases for the exam, proved to be feasible and effective. Furthermore, the cost of developing and administering a new station per student was found to be comparable to the one calculated with other certification examinations reviewed above. These findings further reaffirmed the exam feasibility and cost-efficacy.

Preliminary results suggested that the scores derived from the SCE demonstrated a good level of test sensitivity, construct validity and reliability. Not withstanding the fact that the results need to be further replicated with more candidates and examiners, the present study showed that the SCE scores appeared to be complementary with one another as well as with those derived from the SOE [3] and the Federal final written MC examination. As such, the SCE scores seemed to provide a comprehensive and effective evaluation of the candidates' clinical competency. The reliabilities of the SCE 8-patient station scores were equivalent and as good as those obtained with the ECFMG and NBME 12-patient case examination [6, 22, 23].

However, being a pilot project, this study has the following built-in limitations. First, because all candidates volunteered, they might represent a biased sample and might not be representative of the typical candidates who will take the SCE. However, it is hoped that with the representation of the five medical faculties by the candidates and with the wide range of performance observed among the candidates, this expected bias might be lessened. Second, given the present context concerning the validity of the faculty's involvement in terms of development time of the Federal examination multiple-choice examinations, the reimbursement of the faculty for their participation in the SCE project may introduce a positive bias regarding the faculty's high level of participation in the SCE. It is anticipated that if the SCE is introduced as one of the components of the Swiss final certification examination, a scheme for getting the faculty involved in the SCE elaboration and development would need to be developed. Finally, because this was not a "high-stake" examination but a one-time test limited to a small number of candidates, certain other effects have been observed had the examination been "high-stake" and applied to a larger group of candidates over a longer period. For example, issues of test confidentiality, maintaining equivalence of test administration over days of testing and across testing sites, and keeping a good pool of standardised patients are some of the issues which need to be considered if the exam is to be administered as a large scale "high-stake" examination.

In conclusion, while the preliminary results provide some reassuring evidence on the test feasibility, validity and reliability, additional studies and analyses are needed before the SCE can be formally introduced as part of the Swiss Final Federal examination. Much needed are studies to confirm further the SCE content, structure and scoring validity and the equivalence of the SCE French and German versions, to establish the utility of the SCE exam in providing information on the students' level of clinical performance, to determine and validate the process of setting SCE case and exam passing standards, to develop an SCE model for score reporting and to investigate on its utility for the candidates' application for residency training.

Acknowledgements

We would like to thank Prof. Ralph Bloch and Dr. Peter Frey from the Institute für Aus-, Weiter- und Fortbildung, University of Bern Faculty of Medicine for assisting in the administration of the pilot examination in Bern.

Members of the Pilot federal final examination planning and coordination group – Group co-coordinators: Prof. Ralph Bloch and Mrs. Ursula Hottinger from Bern and Prof. Nu V. Vu from Geneva. Group members: Dr. Wolfgang Gerke and Dr. Christian Schirlo from Zurich, Dr. Gabriele Voigt from Basle, Dr. Martin Perrig from Berne, Dr. Raphael Bonvin from Lausanne, and Dr. Philippe Huber and Dr. Anne Baroffio from Geneva

Case authors and members of the Pilot examination blueprint working group: Dr. Eric Antonelli, Geneva, Dr. Anne Baroffio, Geneva, Dr. Barbara Broers, Geneva, Dr. Christoph Berger, Zurich, Dr. Léo Bühler, Geneva, Dr. Alessandra Canuto, Geneva, Dr. Peter Diem, Bern, Dr. Annick Galetto, Geneva, Dr. Margaret Gerbase, Geneva, Prof. Alain Gervaix, Geneva, Dr. Philippe Huber, Geneva, Dr. Ute Hock, Zurich, Dr. Jean-Paul Humair, Geneva, Prof. Vincenz Im Hof, Bern, Prof. Olivier Irion, Geneva, Mme Carine Layat, Geneva, Prof. Christoph Meier, Geneva, Dr. Mathieu Nendaz, Geneva, Prof. Laurent Nicod, Bern, Dr. Jorg Salomon, Bern, Dr. Erich Seifritz, Bern, Dr. Johannes Streffer, Zurich, Prof. Nu V. Vu, Geneva, Dr. Stephan Wildi, Bern.

Correspondence:
Prof. Nu V. Vu, PhD
Director
University of Geneva Faculty of Medicine
Unit of development and research
in medical education
1 Michel Servet, CMU
CH-1211 Geneva 4
E-Mail: Nu.Vu@medecine.unige.ch

References

1 http://www.iawf.unibe.ch/slo/
2 http://www.bag.admin.ch/berufe/projektmed/gesetz/f/v_ber_f.pdf, http://www.bag.admin.ch/berufe/projektmed/f/index.htm
3 Hottinger U, Krebs R, Hofer R, Feller S, Bloch R. Structured oral examination of the Swiss medical licensing examination: Development and testing in a pilot project. Submitted for publication, 2006.
4 Van der Vleuten CPM, Swanson, DB. Assessment of clinical skills with standardized patients. Teach Learn Med 1990;2:58–76.
5 Vu NV, Barrows HS. Use of Standardized Patients in Clinical Assessments: Recent Developments and Measurement Findings. Educational Researcher 1994;23:23.
6 National Board of Medical Examiners. USMLE Step 2 Clinical skills – Clinical skills research http://www.usmle.org/news/CSEres.pdf
7 Association of American Medical Colleges. Emerging trends in the use of standardized patients. Contemporary Issues in Medical Education 1998;1:1–2.
8 Harden RM, Steven M, Downie WW, Wilson GV. Assessment of clinical competence using objective structure examination. Br Med J 1975;1:447–51.
9 Barrows HS. Simulated (Standardized) patients and other human simulations: A comprehensive guide to their training and use in teaching and evaluation. Chapel Hill, North Carolina: Health Sciences Consortium, 1987.
10 Grand'Maison P, Lescop J, Rainsberry P, Brailowsky CA. Large-scale use of an objective, structured clinical examination for licensing family physicians. Can Med Assoc J 1992;146:1735.
11 Brailovsky, CA, Grand'Maison, P, Lescop, J. A large-scale multicenter objective structured clinical examination for licensure. Academic Medicine 199;67:S37–S39.
12 Tamblyn R, Abrahamowicz M, Brailowsky CA, Grand'Maison P, Lescop J, Norcini J, et al. Association Between Licensing Examination Scores and Resource Use and Quality of care in Primary Care Practice. JAMA 1998;280:989–96.
13 Tamblyn R, Abrahamowicz M, Dauphinee WD, Hanley JA, Norcini J, Girard N, Grand'Maison P, Brailovsky C. Association between licensure examination scores and practice in primary care. JAMA 2002;288:3019–26.
14 Reznick R, Smee S, Rothman A, Chalmers A, Swanson D, Dufresne L, Lacombe G, Baumber J, Poldre P, Levasseur L, et al. An objective structured clinical examination for the licentiate: report of the pilot project of the Medical Council of Canada. Acad Med 1992;67:487–94.
15 Dauphinee, WD, Blackmore, DE, Smee, SM, Rothman, AI, Reznick, RK. Using the judgments of physician examiners in setting the standards for a national multi-center high stakes OSCE. Advances in Health Sciences Education: Theory and Practice. 1997;2:201–11.
16 Smee SM, Dauphinee WD, Blackmore DE, Rothman AI, Reznick RK, Desmarchais JE. A sequenced OSCE for licensure: Administrative issues, results and myths. Advances in Health Sciences Education: Theory and Practice. 2003;8:223–36.
17 Muller ES, Hark P, Margolis M, Clauser B, McKinley D, Boulet JP. An examination of the relationship between clinical skills examination performance and performance on USMLE Step 2. Acad Med 2003;78:S27–S29.
18 Boulet JR, Mc Kinley DW, Norcini, JJ, Whelan GP. Assessing the comparability of standardized patient and physician evaluations of clinical skills. Advances in Health Sciences Education; 2002;7:85–97.
19 Boulet JR, Friedman MBD, Ziv A, Burdick WP, Curtis M, Peitzman S, Gary NE. Using standardized patients to assess the interpersonal skills of physicians. Acad Med 1998;73:S94–S96.
20 Whelan GP, McKinley DW, Boulet JR, Macrae J, Kamholz S. Validation of the doctor-patient communication component of the Educational Commission for Foreign Medical Graduates clinical skills assessment. Med Educ 2001;35:757–61.
21 Boulet JR, Friedman MBD, Hambleton RK, et al. An investigation of the sources of measurement error in the post-encounter written scores from standardized patient examination. Advances in Health Sciences Education 1998;3:89–100.
22 Ziv A, Friedman BDM, Sutnick A, Gary NE. Lessons learned from six years of international administrations of the ECFMG's SP-based clinical skills assessment. Acad Med 1998;73:84–91.
23 Margolis MJ, Clauser BE, Swanson DB, Boulet JR. Analysis of the relationship between score components on a standardized patient clinical skills examination. Acad Med 2003;78:S68–S71.
24 De Champlain AF, Margolis MJ, King A, Klass DJ. Standardized patients' accuracy in recording examinees' behaviours using checklists. Acad Med 1997;72:S85–S97.
25 Margolis MJ, De Champlain AF, Klass JD. Setting examination-level standards for a performance-based assessment of physicians' clinical skills. Acad Med 1998;73:S114–S116.
26 De Champlain AE, MacMillan MK, King A, Klass DJ, Margolis MJ. Assessing the impacts of intra-site and inter-site checklist recording discrepancies on the reliability of scores obtained in a nationally administered standardized patient examination. Acad Med 1999;74:52S–54S.
27 http://www.usmle.org/news/cse/cseftresults2503.htm
28 Shatzer JH, Darosa D, Colliver JA, Barkmeier L. Station-length Requirements for Reliable Performance-based Examination Scores. Acad Med 1993;68:224–9.
29 Shatzer JH, Wardrop JL, Williams RG, Hatch TF. Generalizability of Performance on Different Station-Length Standardized Patient Cases. Teaching and Learning in Medicine 1994; 6:54–8.
30 Chambers KA, Boulet JR, Gary NE. The management of patient-encounter time in a high-stakes assessment using standardized patients. Medical Education 2000;34:813–7.
31 Brailovsky CA, Grand'Maison P. Using evidence to improve evaluation: A comprehensive psychometric assessment of a SP-based OSCE licensing examination. Advances in Health Sciences Education 2000;5:297-219.
32 Papadakis MA, Hodgson CA, Teherani A, Kohatsu ND. Unprofessional behavior in medical school is associated with subsequent disciplinary action by a state medical board. Acad Med 2004;79:244–9.
33 Elstein AS, Shulman LS, Sprafka SA. Medical problem-solving: An analysis of clinical reasoning. Cambridge, Massachusetts: Harvard University Press, 1978.
34 Regehr G, Freeman R, Hodges B, Russell L. Assessing the generalizability of OSCE Measures across Content Domains. Academic Medicine 1999;74:1320–2.
35 Hodges B, McNaughton N, Regehr G, Tiberius R, Hanson M. The challenge of creating new OSCE measures to capture the characteristics of expertise. Medical Education 2002;36:742–8.
36 Swanson DB, Clauser BE, Case SM. Clinical Skills Assessment with Standardized Patients in High-Stakes Tests: A Framework for Thinking about Score Precision, Equating and Security. Advances in Health Sciences Education 1999;4:67.

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

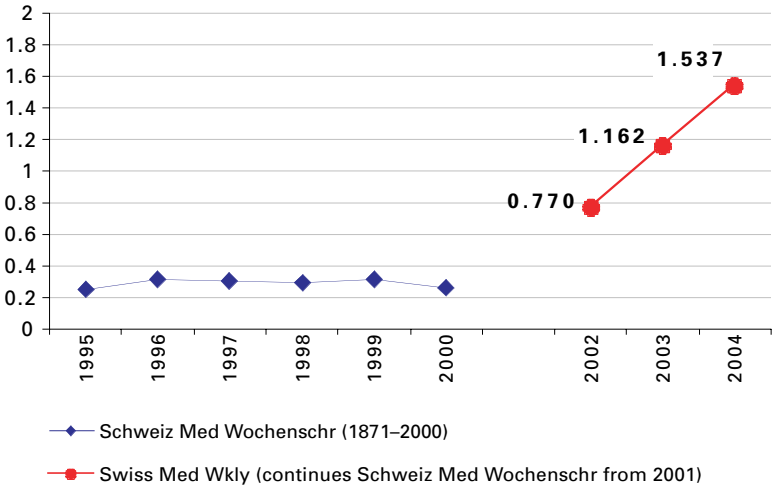
- Fast peer review, by e-mail exchange with the referees
- Prompt decisions based on weekly conferences of the Editorial Board
- Prompt notification on the status of your manuscript by e-mail
- Professional English copy editing
- No page charges and attractive colour offprints at no extra cost

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