A standardized, structured long-case examination of clinical competence of senior medical students

7 authors, including:

Luiz E A Troncon
University of São Paulo
189 PUBLICATIONS **2,023** CITATIONS

Roberto Dantas
University of São Paulo
267 PUBLICATIONS **4,571** CITATIONS

Eduardo Ferriolli
University of São Paulo
125 PUBLICATIONS **1,192** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Improvement of the quality of life of the elderly by the early diagnosis of sarcopenia View project

Project   Consenso Brasileiro de Fragilidade View project

A standardized, structured long-case examination of clinical competence of senior medical students

LUIZ E.A. TRONCON, ROBERTO O. DANTAS, JOSÉ FERNANDO C. FIGUEIREDO, EDUARDO FERRIOLLI, JÚLIO C. MORIGUTI, ANA L.C. MARTINELLI & JÚLIO C. VOLTARELLI

Departamento de Clínica Médica, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Brazil

SUMMARY *Aiming at improving the assessment of senior medical students a standardized and structured modification to the traditional long-case examination is proposed. Students are presented with a sequence of two long cases, with each case being observed by a different examiner. After approaching the patient, the student is interviewed by the examiner who uses a set of four standardized questions. A 10-item checklist covering five groups of major clinical skills was delineated. The time frames for both patient approach (25 minutes) and student interview (10 minutes) were standardized. After the examination, immediate feedback is given to the student. Preliminary results obtained with the first 27 students taking the exam showed that none has failed and 20 of them had combined marks higher than 8.0 (range: 6.6–10.0). Failing performances on individual checklist items were more common in physical examination and less frequent regarding communication with patients. Agreement between different examiners on the assessment of the various skills of the same student was 89%. It is concluded that modifying the format to the long-case examination increased its value in the assessment of student clinical competence.*

Introduction

Assessment of clinical competence of undergraduate medical students plays a key role in their education (Lowry, 1993). However, the application of valid, reliable and feasible examination methods represents a continuing challenge for medical educators (Newble, 1992).

Clerkship in Medicine is an important step in the acquisition of knowledge and clinical skills (McLeod & Harden, 1985). In Brazilian medical schools, assessment of clerks is usually based on the perceptions of supervisors, as there is no tradition of performing objective examination of clinical competence (Troncon *et al.*, 1994).

At the Faculty of Medicine of Ribeirão Preto (University of São Paulo, Brazil), recent changes in the undergraduate curriculum leading to modifications in clerkship configuration provided an opportunity to revise the assessment of senior students. Faculty members in charge of organizing the clerkship in Internal Medicine agreed that a more objective, patient-centered examination was needed for a more valid and reliable summative assessment of the clerks.

The Department of Clinical Medicine has a small amount of experience in running an OSCE (Objective Structured Clinical Examination) (Harden & Gleeson, 1979) for the assessment of basic clinical skills of junior medical students (Troncon *et al.*, 1996). Nevertheless, it was felt that an

examination with a strict OSCE format would not provide comprehensive information on the clinical competence that interns should exhibit in the patient encounter.

We therefore proposed and implemented an examination method containing some elements of both the 'observed long case' (Newble, 1991) and Gleeson's OSLER (Objective Structured Long Examination Record) (Gleeson, 1994, 1977), and running in an OSCE format (Harden & Gleeson, 1979). A main feature of the proposed examination is presenting the student with a sequence of two long cases, each observed by a separate examiner.

In this paper, we describe this method and present the preliminary results of the assessment of a small group of clerks as well as the estimated rate of agreement between different examiners.

Methods

Settings

At the Faculty of Medicine of Ribeirão Preto, students aged 17–19 years enter the medical school shortly after finishing the second cycle of basic education. The current medical school curriculum comprises 2 years of integrated basic sciences, one semester (3rd year) of pre-clinical disciplines and 3 semesters (3rd and 4th years) of clinical disciplines. The two final years (5th and 6th) are spent mainly in five clerkships of Internal Medicine, Pediatrics, Gynecology & Obstetrics, Surgery & Orthopedics and Social & Community Medicine. Each clerkship period lasts 8 weeks per year. During the Internal Medicine periods, students typically engage in intensive ward work in the mornings and in outpatient visits in the afternoons. While in the wards, clerks have considerable responsibility for the management of inpatients. Overnight calls in the Emergency room (1–2 weekly) complete the program.

Examination design

The proposed examination consisted of a sequence of two long cases entirely observed and organized in a multiple-station format. The time frames for both student–patient encounter and examiner–student interview are standardized. Students have 25 minutes with the first patient (focused

Correspondence: Dr Luiz E.A. Troncon, Departamento de Clínica Médica, Hospital das Clínicas - Campus da USP, 14048-900, Ribeirão Preto, SP, Brazil. Tel: +55 16 602 2457; fax: +55 16 633 6695; email: ledatron@fmrp.usp.br

history taking and physical examination), after which 2 minutes are allowed for data organization and preparation for the interview with the examiner. This takes the next 8 minutes and is structured over four standardized questions aimed at assessing clinical acumen, clinical reasoning and patient management. The interview is followed by the provision of immediate feedback to the student on his/her performance. The student then moves to the next station, where the same sequence is repeated with another patient and examiner.

Performance assessment

Student performance on both patient approach and interview is assessed and graded using a 10-item checklist covering five groups of clinical skills: (a) history-taking; (b) physical examination; (c) communication and interaction with the patient; (d) case presentation and clinical reasoning and (e) patient management (Figure 1). The last four checklist items are covered during the interview, which is carried out over a set of four standard questions (Figure 2). The checklist also allows for the evaluation of case difficulty during actual student assessment. Since sources of difficulty may relate to different domains, the checklist contains grades (low, average and high) for three aspects: patient ability to communicate, physical examination findings and clinical problem as a whole.

Grading and marking

For each of the 10 checklist items, a four-grade marking system is applied: (a) fully adequate (mark 1.0); (b) barely adequate (mark 0.75); (c) inadequate (mark 0.25) and (d) poor (mark 0). An intermediate grade corresponding to a mark 0.5 was deliberately omitted in order to avoid borderline evaluations. A detailed description of student behaviors to be associated with each grade for all items was also constructed. For example a 'fully adequate' grade on 'characterization of the main complaint' should be attributed to students who 'obtained information on type, intensity, chronology and modifying circumstances of the chief complaint presented by the patient, as well as on the influence of the present disease on patient's ordinary activities and general well being'.

Preparation for the examination

The preparation for the examination involved a number of meetings between faculty members and student representatives aimed at reaching consensus on the main points of the examination. Case characteristics and degree of difficulty, as well as the range of specific contents and clinical problems to be included in the examination, were carefully discussed. Those faculty who would probably act as examiners received detailed written instructions for the entire procedure and were put in charge of selecting and recruiting real patients to be included in the exam. The criteria for selection included stable clinical conditions, absence of communication problems, presence of at least one well-defined diagnostic abnormality detectable by a standard physical examination maneuver, well-defined diagnosis fitting into the previously elaborated list of contents and problems and an average degree of overall difficulty. All recruited patients had also to give their informed consent.

Running the examination

The actual run-up to the examination took place in a special ward unit containing six rooms located around a living room. Each room was arranged as an office and accommodated one patient. For each run of the examination, a pool of 9–12 patients was selected, which was sufficient for the patients to alternate between working with students (25 min) and resting (15–50 min). Each selected patient was approached by students from two to four times during the examination day. Each station was staffed by one member of faculty, who was fully aware of the main findings and clinical problems of the patients alternating in that particular station. Immediately before the examination, both examiners and students were briefed on the main points of the exam. Clerks were examined in groups of six. Since each run took 40 minutes and every student was examined in two consecutive stations, the time needed for the assessment of up to 18 interns was 4 hours.

Data analysis

In the first two runs of the proposed examination a group of 27 students was examined. For every student, the marks given by each examiner to the various checklist items were added to form a 'total mark', which could range from 0 to 10. The average of the total marks attributed by the two examiners comprised the 'combined mark'. In order to comply with local regulations, students had to obtain a combined mark of at least 5.0 to 'pass' the examination.

For the assessment of overall student performance on the various checklist items, any of the two 'adequate' grades were regarded as a 'pass' event and any 'inadequate' or 'poor' grade corresponded to a 'fail' event. The frequency of 'fail' events was calculated as a proportion of 540, the total number of grades attributed (27 students times 10 items times 2 examiners).

Inter-examiner variability was estimated by determining the rates of 'full agreement', 'partial agreement' and 'disagreement' concerning the 10 checklist items, as a proportion of the total number of decisions taken. A 'full agreement' was recorded when both examiners attributed an identical grade for a given item. A 'partial agreement' was recorded when both grades, despite being different, were in the same 'pass' or 'fail' ranges. A 'disagreement' was recorded when grades fell in different ranges ('pass' or 'fail'). Considering that the limited numbers of students examined would preclude an appropriate statistical analysis, no attempt was made to calculate reliability coefficients more accurately.

Results

All 27 students examined carried out the expected tasks and all the 54 checklist forms were thoroughly fulfilled by examiners. Also, all the selected patients complied with the proposed schedule and no incidents or drop-outs were recorded in any of the two runs of the examination.

Student performance

Among the first 27 interns examined, none failed. Combined marks ranged from 6.6 to 10.0 (median: 8.5) and were equal or higher than 8.0 for 20 students.

Practice-based assessment of clerks in internal medicine

Student name:...

Examiner name:...

Date:......./......./...........

Checklist

History taking   **Grading(*)**

1. Characterization of main complaint.....................	(FA)	(BA)	(I)	(P)
2. Characterization of associated manifestations and other personal and familial data...................	(FA)	(BA)	(I)	(P)

Physical examination

3. Approach to the most likely affected organ or system (technique and thoroughness)................	(FA)	(BA)	(I)	P)
4. Routine general examination................................	(FA)	(BA)	(I)	(P)

Communication and interaction with patient

5. Technical aspects of medical interview................	(FA)	(BA)	(I)	(P)
6. Personal respect and consideration......................	(FA)	(BA)	(I)	(P)

Case presentation and clinical reasoning

7. Summary of main points of the case....................	(FA)	(BA)	(I)	(P)
8. Diagnostic hypothesis...	(FA)	(BA)	(I)	(P)

Patient management

9. Indication of diagnostic measures........................	(FA)	(BA)	(I)	(P)
10. Indication of therapeutic measures......................	(FA)	(BA)	(I)	(P)

(*) FA = fully adequate (mark 1.0); BA = barely adequate (0.75); I = inadequate (0.25); P = poor (0.0).

Observations ....

....

....

Case difficulty   **Grading (#)**

A. Patient communication	(H)	(A)	(L)
B. Physical findings	(H)	(A)	(L)
C. Clinical problem	(H)	(A)	(L)

(#) H = high; A = average; L = low.

Figure 1. Assessment form for the proposed examination of clinical competence.

'Fail' performances on any of the checklist items were recorded in only 36 (6.6%) out of the 540 grades attributed. Table 1 shows that 26 of these events were recorded for the groups of skills directly related to patient approach (physical examination and history taking) and only 10 'fail' performances were detected for the items involving cognitive performance (clinical reasoning and patient management).

Case characteristics

The two runs of the proposed examination involved 18 patients. Data regarding case difficulty are presented in

Practice-based assessment of clerks in internal medicine
Standardized questions for clerk interview
1. Would you please make a summary of the main points regarding this case that you gather from the clinical history and the physical examination ?
2. Would you please describe briefly what are your main diagnostic hypotheses, ranked from the most to the least probable?
3. Taking into account your findings, would you please mention briefly what tests you would need to confirm your diagnostic hypothesis?
4. Would you please describe briefly what kind of general and specific therapeutic measures you would take in order to both solve the patient's problem and improve his/her well-being?

Figure 2. Standardized questions for the assessment of clerks' competences on case presentation, clinical reasoning and patient management.

Table 1. Number of student 'fail' performances on the standardized, structured long-case examination.

Clinical skills group	'Fail' performances (%)
History taking	9
Physical examination	15
Communication with patient	2
Case presentation and clinical reasoning	4
Patient management	6

Table 2, which shows that only a small proportion of cases (1.8%) were regarded as having difficulties in communicating with students. Physical examination was considered as having low or average difficulty in 92.6% of cases. As expected from the requirements for inclusion in the examination, most of the cases included (87%) had average or low difficulty concerning the nature of the clinical problem presented.

Agreement between examiners

Data for the rates of agreement between examiners concerning student performance throughout the 10 checklist items are presented in Table 3. In 89% of the decisions taken, there was full or partial agreement between the two examiners and the rate of disagreement was quite low (11.0%).

Discussion

In many medical schools throughout the world, assessment of clinical competence of senior students is carried out by

Table 2. Degree of difficulty of clinical cases included in the observed long-case examination.

Category	Degree of difficulty (%)		
	High	Average	Low
Patient communication	1.8*	9.2	89.2
Physical findings	7.4	18.5	74.1
Clinical problem	13.0	50.0	37.0

Note: *Percentage of total number of decisions taken.

Table 3. Examiners' agreement on student clinical competence.

Category	% of grades attributed
Full agreement*	43.0
Partial agreement#	46.0
Disagreement$^\phi$	11.0

Notes: *both examiners attribute identical grades;
examiners attribute different grades but in the same range ('pass' or 'fail'); $^\phi$examiners attribute grades in different ranges.

the traditional long-case examination (Stokes, 1974; Weatherall, 1991; Newble, 1992; Lowry, 1993), which has been criticized for its unreliability. Nevertheless, the place of the long-case examination in the assessment of students seems to be justifiable not only for its tradition and practicality, but also because this method has an adequate degree of face validity (Newble, 1991; 1992; Gleeson, 1997).

Modifications of the traditional long case (Newble, 1992; Price & Byrne, 1994; Abouna & Hamdy, 1999), including the OSLER (Gleeson, 1994, 1997) have improved the method, but it still has a number of shortcomings. For example, in the 'observed long case' (Newble, 1991) the checklist seems to be too generic to guarantee that individual components of competence are properly assessed. On the other hand, in Gleeson's OSLER (Gleeson, 1997), students are observed only partially and the examination seems to put much more emphasis on the assessment of case presentation and the ability of the student to communicate with the examiner, rather than on basic clinical skills.

In the long-case examination herein described, the student performance is entirely observed and the focus of the assessment is directed to genuine clinical skills. In fact, six out of the 10 items comprising the checklist are devoted to patient-centered skills. Moreover, we have used real patients presenting with clinical features fitting into a predefined set of contents and clinical problems. Therefore, the actual observation of the student performing relevant clinical tasks, complemented by the use of standard questions for the assessment of the cognitive aspects involved, implies coverage of the most important components of medical competence (Newble, 1992). Thus, our proposed examination is in line with recent work aimed at improving student

evaluation using a more standardized performance-based exam (Abouna & Hamdy, 1999).

Like others (Newble, 1991), we also proposed a sequence of two long cases and a more rigid time control, which were implemented in a two-station OSCE format (Harden & Gleeson, 1979), so as to ensure that all clerks were assessed in similar conditions. While exposing the candidate to two different cases increased the sample of contents (Newble, 1991; Van der Vleuten, 1996), the use of more than one observer is likely to have improved exam accuracy (Van der Vleuten, 1996; Gleeson, 1997; Abouna & Hamdy, 1999).

Unlike others (Newble, 1991, Price & Byrne, 1994, Gleeson, 1994, 1997; Abouna & Handy, 1999), we have not used two simultaneous examiners per case, which was done for a number of practical reasons. The number of examiners available in our department is quite small and students unaccustomed to being observed pointed out that the presence of two examiners would probably make the exam more stressful. Also, faculty expressed concerns about the efficacy of the 'confer and agree' process between concomitant examiners (Gleeson, 1997) on 10 checklist items within a controlled time frame. Also taken into consideration was the opinion of international experts, who think that presenting the candidate with two cases, each one with a different single examiner, is likely to improve reliability of the long-case examination (Van der Vleuten, 1996). The low rate of disagreement between examiners (11%) that we found in our long-case examination is consistent with this view.

The use of one examiner per station also seemed to increase practicality, by means of a more convenient usage of the human resources available. Indeed, the organizational difficulties of running this examination were felt to be lower than those of a conventional OSCE for assessing basic clinical skills of junior medical students (Troncon *et al.*, 1996). The examination of an entire group of interns demanded six examiners, for one morning's work, which was perceived as being reasonable in terms of our manpower utilization.

Our preliminary results showed that not a single intern failed and the combined marks were rather higher than one could predict. A possible explanation for this could be that examiners were reluctant to attribute grades in the 'fail' range, since nearly all examiners were clerkship supervisors and involve intensively with students in their daily work on the wards. A possible way to overcome this limitation would be inviting external examiners (Walters *et al.*, 1995), which would probably increase exam accuracy. Nevertheless, this would probably work against the feasibility of the proposed examination.

As far as the overall student performance on individual items is concerned, we found a relatively high frequency of deficiencies regarding physical examination, which cannot be explained by difficulties expressed by the selected cases (Table 2). This is consistent with data reported elsewhere (Gleeson, 1994), which has been interpreted as a possible consequence of a gradual decline in the time available for teaching clinical skills in the undergraduate curriculum (Gleeson, 1994).

Provision of feedback to students immediately after the examination has been recognized as a substantial improvement in the educational value of current methods of assess-

ment of clinical competence (Newble, 1992; Lowry, 1993; Gleeson, 1994, 1997; Van der Vleuten, 1996). Although we have not attempted to evaluate the response of students to the proposed examination, many clerks expressed informally their satisfaction concerning this particular point of the exam. We also believe that the involvement of students in the delineation of the assessment has contributed to both increase student compliance with the implementation of the exam and improve its educational meaning.

Acknowledgement

Notes on contributors

Luiz E.A. Troncon, MD, is an associate professor in the Department of Clinical Medicine and former Chairman of the Committee of Undergraduate Studies, Ribeirão Preto Faculty of Medicine, University of São Paulo, Brazil.

Roberto O. Dantas, MD, is an associate professor in the Department of Clinical Medicine and former Coordinator of the Internship in Internal Medicine, Ribeirão Preto Faculty of Medicine, University of São Paulo, Brazil.

Jose Fernando C. Figueiredo, MD, is an assistant professor in the Department of Clinical Medicine, Ribeirão Preto Faculty of Medicine, University of São Paulo, Brazil.

Eduardo Ferrioli, MD, is an assistant physician and clerk supervisor in the Department of Clinical Medicine, Ribeirão Preto Faculty of Medicine, University of São Paulo, Brazil.

Julio C. Moriguti, MD, is an assistant physician and clerk supervisor in the Department of Clinical Medicine, Ribeirão Preto Faculty of Medicine, University of São Paulo, Brazil.

Ana L. C. Martinelli, MD, is an assistant professor in the Department of Clinical Medicine, Ribeirão Preto Faculty of Medicine, University of São Paulo, Brazil.

Julio C. Voltarelli, MD, is an associate professor in the Department of Clinical Medicine, Ribeirão Preto Faculty of Medicine, University of São Paulo, Brazil.

References

Abouna, G.M. & Hamdy, H. (1999) The integrated direct observation clinical encounter examination (IDOCEE)—an objective assessment of students' clinical competence in a problem-based learning curriculum. *Medical Teacher*, 21, pp. 67–72.

Gleeson, F. (1994) The effect of immediate feedback on clinical skills using the OSLER, in: A.I. Rothman & R. Cohen (Eds) *Proceedings of the Sixth Ottawa Conference on Medical Education*, pp. 412–415 (Toronto, University of Toronto Bookstore Custom Publishing).

Gleeson, F. (1997) *AMEE Education Guide No. 9: Assessment of Clinical Competence using the Objective Structured Long Examination Record (OSLER)* (Dundee, Association for Medical Education in Europe).

Harden, R.M. & Gleeson, F.A. (1979) Assessment of clinical competence using an objective structured clinical examination (OSCE), *Medical Education*, 13, pp. 39–45.

Lowry, S. (1993) Assessment of students, *British Medical Journal*, 306, pp. 51–54.

McLeod, P.J. & Harden, R.M. (1985) Clinical teaching strategies for clinicians, *Medical Teacher*, 7, pp. 173–189.

Newble, D.I. (1991) The observed long case in clinical assessment, *Medical Education*, 25, pp. 369–373.

Newble, D.I. (1992) Assessing clinical competence at the undergraduate level, *Medical Education*, 26, pp. 504–511.

PRICE, J. & BYRNE, G.J.A. (1994) The direct clinical examination: an alternative method for the assessment of clinical psychiatry skills in undergraduate medical students, *Medical Education*, 28, pp. 120–125.

STOKES, J. (1974). *ASME Medical Education Booklet No.2:The Clinical Examination: Assessment of Clinical Skills* (Dundee, Association for the Study of Medical Education).

TRONCON, L.E.A., FOSS, N.T., VOLTARELLI, J.C. & DANTAS, R.O. (1996) Avaliação de habilidades clinicas por exame objetivo estruturado por estações, com emprego de paciente padronizados: uma aplicação no Brasil (Parte II), *Revista Brasileira de Educação Médica*, 20, pp. 53–60.

TRONCON, L.E.A., RODRIGUES, M.L.V., PICCINATO, C.E. & CIANFLONE, A.R.L. (1994) Responses of students and faculty members to changes in assessment in a Brazilian medical school, in: A.I. ROTHMAN & R. COHEN (Eds) *Proceedings of the Sixth Ottawa Conference on Medical Education*, pp. 126–131 (Toronto, University of Toronto Bookstore Custom Publishing).

VAN DER VLEUTEN, C. (1996) Making the best of the long case, *Lancet*, 347, pp. 704–705.

WALTERS, W.A., SIVANESARATNAM, V. & HAMILTON, J.D. (1995) External examiners, *Lancet*, 345, pp. 1093–1095.

WEATHERHALL, D.J. (1991) Examining undergraduate examinations, *Lancet*, 338, pp. 37–39.