University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Sutherland, RM;Reid, KJ;Chiavaroli, NG;Smallwood, D;McColl, GJ

Title:
Assessing Diagnostic Reasoning Using a Standardized Case-Based Discussion

Date:
2019-05-20

Citation:
Sutherland, R. M., Reid, K. J., Chiavaroli, N. G., Smallwood, D. & McColl, G. J. (2019). Assessing Diagnostic Reasoning Using a Standardized Case-Based Discussion. JOURNAL OF MEDICAL EDUCATION AND CURRICULAR DEVELOPMENT, 6, https://doi.org/10.1177/2382120519849411.

Persistent Link:
https://hdl.handle.net/11343/251214

License:
CC BY-NC

Assessing Diagnostic Reasoning Using a Standardized Case-Based Discussion

Journal of Medical Education and Curricular Development
Volume 6: 1–6
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2382120519849411

Ruth M Sutherland[1], Katharine J Reid[1], Neville G Chiavaroli[1], David Smallwood[2] and Geoffrey J McColl[3]

[1]Department of Medical Education, Melbourne Medical School, The University of Melbourne, Melbourne, VIC, Australia. [2]Royal Melbourne Clinical School, Melbourne Medical School, The University of Melbourne, Melbourne, VIC, Australia. [3]Melbourne Medical School, The University of Melbourne, Melbourne, VIC, Australia.

ABSTRACT

BACKGROUND: Development of diagnostic reasoning (DR) is fundamental to medical students' training, but assessing DR is challenging. Several written assessments focus on DR but lack the ability to dynamically assess DR. Oral assessment formats have strengths but have largely lost favour due to concerns about low reliability and lack of standardization. Medical schools and specialist medical colleges value many forms of oral assessment (eg, long case, Objective Structured Clinical Examination [OSCE], viva voce) but are increasingly searching for ways in which to standardize these formats. We sought to develop and trial a Standardized Case-Based Discussion (SCBD), a highly standardized and interactive oral assessment of DR.

METHODS: Two initial cohorts of medical students (n = 319 and n = 342) participated in the SCBD as part of their assessments. All students watch a video trigger (based on an authentic clinical case) and discuss their DR with an examiner for 15 minutes. Examiners probe students' DR and assess how students respond to new standardized clinical information. An online examiner training module clearly articulates expected student performance standards. We used student achievement and student and examiner perceptions to gauge the performance of this new assessment form over 2 implementation years.

RESULTS: The SCBD was feasible to implement for a large student cohort and was acceptable to students and examiners. Most students and all examiners agreed that the SCBD discussion provided useful information on students' DR. The assessment had acceptable internal consistency, and the associations with other assessment formats were small and positive, suggesting that the SCBD measures a related, yet novel construct.

CONCLUSIONS: Rigorous, standardized oral assessments have a place in a programme of assessment in initial medical training because they provide opportunities to explore DR that are limited in other formats. We plan to incorporate an SCBD into our clinical assessments for the first year of clinical training, where teaching and assessing basic DR is emphasized. We will also explore further examiners' understanding of and approach to assessing DR.

KEYWORDS: diagnostic reasoning, clinical reasoning, oral assessment, Standardized Case-Based Discussion, medical education

RECEIVED: March 8, 2019. **ACCEPTED:** April 16, 2019.

TYPE: Methodology

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Katharine J Reid, Department of Medical Education, Melbourne Medical School, The University of Melbourne, Melbourne, VIC 3010, Australia. Email: kjreid@unimelb.edu.au

Introduction

Accurate diagnostic reasoning (DR) is fundamental to ensuring patient care and safety, and thus the development of DR is a key component of medical training. Valid and reliable assessment methods are central to evaluating medical students' DR skills. However, there is no consensus on the most effective approaches to evaluating these reasoning skills. Currently, several written assessments aim to assess DR, including the Script Concordance Test,[1] Key Features Test,[2] Clinical Reasoning Problems,[3] and the Extended Matching Format.[4] Although each of these assessments may capture some aspect of DR, the written format cannot capture the dynamic nature of DR in practice, such as changing reasoning in response to new information or explaining and defending one's reasoning.

Oral assessments have been widely used to evaluate reasoning skills,[5] but face criticism because of lack of standardization, risk of bias, and poor reliability.[6] For example, traditional long cases provide extended assessment time for students to formulate and convey their understanding of a real patient's problem.[7] Long case advocates value its authenticity; yet, the long case has also been subject to criticism of its lack of standardization due to examiner and patient variability. Such concerns have seen oral assessments lose favour, including for assessment in clinical education contexts.[8] It is possible to introduce, as we have done, direct questions about students' DR into the Objective Structured Clinical Examination (OSCE), but time constraints limit extended exploration of DR in this setting. However, avoiding oral assessments altogether appears at odds with programmatic

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

assessment, which gathers multiple authentic samples of student performance over time using different assessment methods linked to a variety of consequences.[9] Ignoring oral assessment formats significantly compromises a detailed and nuanced understanding of students' DR development.

Appropriately designed and implemented oral assessments allow educators to clarify, probe, and confirm students' reasoning, while also reflecting the reality of clinical learning in the workplace. Arguably, for a complex and contextual skill such as DR, it may be the most appropriate method of assessing medical students' performance. Therefore, from a pedagogical perspective, we argue that it is desirable to use oral assessments to advantage while minimizing their weaknesses. We therefore developed a new standardized, oral assessment approach specifically to investigate students' DR. Across a 2-year implementation, we sought to gather data on the performance of the Standardized Case-Based Discussion (SCBD; its score distributions, internal consistency, and relationship to other assessment components) and to determine whether SCBD examiners and students regarded the assessment as an acceptable approach to assessing DR. Our aim was to judge whether the SCBD had value as an ongoing component of our assessment programme.

Method

Development of the SCBD

We developed and implemented a single SCBD as a component of the assessment of the third year (the second year of full-time clinical training) of a 4-year postgraduate medical course. SCBD comprises a face-to-face oral assessment between student and examiner, preceded by a video trigger of a doctor interviewing a patient about a new clinical problem. Our implementation design for the SCBD involved development, piloting, improvements based on feedback, trial implementation in 2015, and final implementation in 2016. A single case required substantial investment of resources at the development stage (writing, reviewing, filming, preparing stimulus materials). As a result, it is feasible in the context of limited time and financial resources to produce only 1 case per year. Our trade-off was to produce cases with a high degree of standardization and authenticity and to only allocate the case a 10% weighting in the aggregate subject mark to balance the possibility of low reliability of a single case.

SCBD administration comprises 2 phases. In phase 1 (15 minutes), students watch a 5-minute video trigger on an individual computer, with headphones, in a computer laboratory. All students receive the same patient information, may replay the video, and take notes. In phase 2 (15 minutes), students discuss the case, request additional relevant clinical information, and justify their DR with a trained clinical examiner.

Development of the SCBD involved scripting and filming the video trigger, developing comprehensive clinical information related to the patient presentation, producing a marking sheet to guide examiners through the history, examination, investigations, and management phases, and a marking rubric for defining the expectations for students' performance. A panel of clinical experts reviewed a first draft of these elements of the SCBD to ensure accuracy and realism. The panel reviewed the materials critically and provided detailed feedback to the authors on aspects which the group agreed should be revised. To further standardize examiner marking, the expert panel also created a document outlining the performance expectations for a good and a borderline student at this stage of medical training.

We chose a video instead of a paper-based trigger to increase the assessment's authenticity, to mimic real patient encounters, and because clinicians obtain valuable information from patients' non-verbal and verbal cues. The cases were undifferentiated patient presentations based on real patients (eg, a young woman with abdominal pain, an undifferentiated dyspnoea in an older patient). Generalists developed the cases and trigger scripts, to maintain an undifferentiated feel to the cases, thereby allowing several diagnostic possibilities to be explored. The presenting complaints of the patients in the trigger videos were designed to be common complaints that the students would have been exposed to many times throughout their training and were not symptoms limited to 1 organ system (eg, dyspnoea can be a presenting complaint for cardiovascular, respiratory, haematological, and psychological conditions). SCBD trigger information was intentionally broad, so that no specific illness script was triggered, and the hypothetico-deductive method was required to further explore diagnostic possibilities. Fundamentally, the assessment required examiners to make an informed judgement about the capacity of the student to think critically about the case, through applying their biomedical knowledge and weighing up the available clinical information and additional investigations.

In total, 5 clinician examiners and 12 volunteer final-year medical students trialled a pilot case to refine the SCBD before implementation in the assessment programme. Examiners and students provided feedback on timing, materials, and examiner interaction, during separate group interviews. Overall, students and examiners regarded the SCBD pilot favourably. Minor suggestions for improvement prior to implementation included adjusting the timing of warning bells and developing more specific guidelines for probing students' reasoning.

Students prepared for the initial implementation of the SCBD by watching an example video trigger and a simulated example of a good student performance on their Learning Management System. Our SCBD design involved a clinical conversation between a student and a single examiner (volunteer clinicians who are staff members at the university or affiliated clinical schools). Without a second rater, we established several measures to try to standardize examiner behaviour. Examiners prepared by completing the mandatory SCBD online training module. This module outlined the aims and

process of the SCBD and included a calibration exercise where examiners watched and assessed a simulated student performance (at a borderline and a good level) and compared their ratings with those of an expert panel. Examiners were trained to 'probe' students' reasoning as they articulated it, without prompting them for particular points. The video trigger and clinical information documents were distributed to examiners before the examination. On the day of the examination, examiners reviewed the video trigger as a group to reinforce the online training and to allow for case-specific questions.

All SCBD were completed in one half-day for more than 300 students (approximately 35 examiners work across the morning, examining individually in about 30 separate classrooms, examining up to 10 students across the morning). Examiners initiated the discussions by asking students to list their differential diagnoses and to justify their reasoning. To guide examiners to appropriately probe students' reasoning, set questions were included on the mark sheet at the beginning and end of each section, for instance, 'What further specific information would you like to know about the presenting complaint?' and 'How does this information help?' Examiners provided further clinical information to the students on request. Students were responsible for managing their time, although examiners understood that they could ask for the student's consent to move forward, if they believed that the student had exhausted their current line of reasoning or had satisfactorily justified their reasoning. To assist the students' time management, there was a warning bell at 10 minutes with a final bell at 15 minutes.

Examiners marked students' reasoning about the patient's history and how this informed their thinking about likely diagnoses, requested examination findings, investigations requested and their interpretation, and the proposed patient management sections of the SCBD on a 5-point scale where 0 = *unsatisfactory* and 4 = *excellent*, and rated students' overall performance using a 5-point global scale (where 1 = *fail* and 5 = *excellent*) for standard setting purposes. Examiners' use of these scales was guided by a written rubric which included 4 to 7 descriptors of the characteristics of performance at each level of the scale. Students' total score (out of 20) was the sum of the sections with history double-weighted. Double weighting reflected the high importance of medical history to diagnosis[10] and the expectation that more time would be allocated to this phase. This weighting also approximately reflected the weighting in the third-year curriculum and learning objectives.

Students and examiners completed a survey at the end of the assessment period that asked for feedback on their experience of the SCBD related to processes, acceptability, and value as an assessment. These data comprised quantitative Likert-type scale items on the student and examiner surveys as well as open-ended comments on the student survey. The data related to examining the performance of the SCBD and evaluation data from students and examiners about their experience of the SCBD were gathered as part of quality assurance and did not require institutional ethics approval but are conducted under the oversight of the medical course evaluation committee. All students are informed at the beginning of each year that assessment and evaluation data may be used anonymously for the purposes of quality assurance and curriculum development.

Table 1. Descriptive statistics for the SCBD.

	IMPLEMENTATION YEAR	
	2015	2016
Mean	14.6	12.5
SD	3.95	3.72
Median	15	12
Min	0	2
Max	20	20
Cronbach alpha	0.82	0.77
No. of students	319	342
No. of examiners	35	35

Abbreviations: SCBD, Standardized Case-Based Discussion.

Data analysis

The trial and implementation of the SCBD yielded data related to student achievement on the SCBD (a total score out of 20), quantitative evaluative data (Likert-type agreement scales where 1 = *strongly disagree* and 5 = *strongly agree*) from students and examiners, and qualitative feedback data from students. The quantitative data on students' total score on the SCBD were analysed descriptively, using means and standard deviations, and correlations with other assessment components and the final subject score (the weighted sum of all assessments in the year), Cronbach alpha was calculated as a measure of internal consistency of the 4 subscores, and the quantitative evaluative data were analysed descriptively, using means and standard deviations. The qualitative data were coded thematically by the second author (K.J.R.) with the themes verified independently by the first author (R.M.S.).

Results

We gathered information on the SCBD over 2 implementation years to provide detailed information on score distributions, internal consistency of the assessment (using Cronbach alpha), relationships between SCBD scores and other assessment forms (both written and clinical), and acceptability of the assessment for students and examiners.

It can be seen from Table 1 that the average raw SCBD scores in the first year were slightly higher than those in the second year, although examiners used the full range of scores in both years. Internal consistency was acceptable on both administrations of the SCBD. Correlations with other assessment formats

Table 2. Mean ratings for the student and examiner evaluation of the Standardized Case-Based Discussion.

STUDENTS	IMPLEMENTATION YEAR			
	2015 (N = 285)		2016 (N = 230)	
REGARDING THE STANDARDIZED CASE-BASED DISCUSSION	MEAN	SD	MEAN	SD
The materials online were helpful in preparing for the assessment	3.4	0.96	3.1	1.09
The video trigger provided me with enough information to begin the assessment	4.0	0.79	3.8	0.89
There was sufficient time to think about the video trigger	3.8	0.97	3.5	1.13
The discussion with the examiner allowed me to demonstrate my clinical reasoning	3.8	1.03	3.5	1.14
I was able to manage my time effectively	3.6	1.05	2.9	1.21
The 10-minute warning bell helped me to manage my time	3.7	0.98	3.3	1.21
EXAMINERS	2015 (N = 32)		2016 (N = 27)	
REGARDING THE STANDARDIZED CASE-BASED DISCUSSION	MEAN	SD	MEAN	SD
This assessment provides useful information on students' diagnostic reasoning	4.5	0.51	4.5	0.51
Students managed their time effectively during the assessment	3.8	0.93	3.1	0.91
The 10-minute warning bell helped students to manage their time effectively	4.0	0.86	3.7	0.91
I was confident in probing students' thinking (without prompting)	4.0	0.86	3.8	0.85
There was sufficient clinical information available to the examiners	4.3	0.62	4.4	0.70
The following documents were helpful in assessing student performance:				
Outline of expectations	4.4	0.49	4.3	0.48
Marking sheet	4.1	0.53	4.1	0.60
Overall rubric	4.1	0.66	4.2	0.70
Examiner training:				
Prepared me for my role as an examiner	4.2	0.48	3.9	0.95
Provided guidance on probing students' thinking	4.0	0.60	3.8	0.96
Provided useful material	4.3	0.51	4.1	0.87
I would be happy to examine SCBD in the future	4.6	0.57	4.5	0.51

Abbreviations: SCBD, Standardized Case-Based Discussion.
Evaluation items measured on a 5-point Likert-type scale where 1 = *strongly disagree* and 5 = *strongly agree*.

in the medical course were relatively consistent across years and ranged from negligible (for Mini-Clinical Evaluation Exercise [Mini-CEX]: 0.07-0.11) to weakly positive for all other assessment types, both written (multiple-choice questions [MCQs]: 0.27-0.37, short-answer questions [SAQs]: 0.29-0.35, International Foundations of Medicine [IFOM]: 0.32-0.34) and clinical (OSCEs: 0.32-0.33). Overall, SCBD had a weak positive relationship with the final subject score (the weighted sum of all assessments in the year) in both years (0.35-0.41). Achievement on the SCBD contributed 10% to students' final mark at the end of their second year of clinical training. We used the borderline regression standard setting method[11] to determine the cut score for satisfactory performance. The SCBD cut score was 8 out of 20 for both cases. To support our investment in a standardized clinical assessment format, we sought further information on the degree of examiner variability in scoring across the two years of implementation. A between-group analysis of variance (ANOVA) showed that, on average, only 17% of the variation in scores could be attributed to variation in scoring between examiners.

Students and examiners evaluated their experience of the SCBD (Table 2). Most students agreed that the discussion with the examiner allowed them to demonstrate their DR. For examiners, 100% in each year agreed that the SCBD provided useful information on students' DR. Students were particularly positive about the pre-assessment video trigger believing that they had sufficient time to think about the video trigger and that it provided sufficient information. In the first year, students were positive about their time management with the assistance of the warning bell. Students were less positive

Table 3. Percentage of students (n = 179) providing comments on the SCBD.

POSITIVE COMMENTS		EXAMPLE
Good assessment form	37.4%	*Really good method of assessment. Allowed us time to reason and have a discussion, much better opportunities than OSCEs provide.* *This was a great form of assessment as it allows us to demonstrate critical thinking rather than memorizing lists of things to ask.*
Opportunity to demonstrate reasoning	24.0%	*Very, very useful assessment method as I was able to demonstrate my knowledge/clinical reasoning which isn't always available in other assessments-best assessment type!!* *I think the SCBD is a good way to demonstrate clinical reasoning and if anything less structured than OSCE. It allows more time to gather your thoughts.*
Preferred over other assessment forms	15.1%	*A fantastic way to assess clinical reasoning. There should be more SCBDs than OSCEs, more applicable for testing skills needed as a junior doctor.*
Fairer assessment form	5.0%	*Fair assessment of diagnostic reasoning*
Other	7.3%	*Very well organized and done assessment, I really liked it.*
NEGATIVE COMMENTS		**EXAMPLE**
Concerns about standardization	20.1%	*My examiner, unlike others, barely prompted me at all throughout the whole discussion for my reasoning, which gave no indication of when and how much reasoning was wanted.* *Some students did not receive as much prompting as others.*
Concerns about time management	12.8%	*I felt it was hard to manage time. We had no idea what to focus most on.*
Contributes too much to final mark	7.8%	*Worth a lot of marks for something so new.*
Inconsistency between preparation and SCBD	5.6%	*The video showed an examiner asking questions but I felt like I wasn't asked many questions as I was expecting from the example.*
Concerns about examiner	5.0%	*Examiner kept going 'what else do you want on exam' I was asked five times which made me think I had forgotten something important-this wasted time and I did not finish.*
Other	4.5%	*Very subjective. Some examiners provided options to choose from regarding management. Leading questions as well. Unfair.*
SUGGESTED IMPROVEMENTS		**EXAMPLE**
Provide opportunities for practice	8.4%	*We need practice for it throughout the year.*
Provide more guidance on expectations	7.8%	*I think we needed to be clearly instructed to reason our decision making during the SCBD. Clearer instruction on marking schema needed to be provided to students earlier.*

Abbreviations: OSCE, Objective Structured Clinical Examination; SCBD, Standardized Case-Based Discussion.
Percentages do not add to 100 because some students provided more than 1 comment.

overall in the second year, particularly on features related to timing. In both years, students were least positive about the online learning resources in preparing for the assessment. Examiner evaluation of the SCBD was very positive. The exceptions were items related to student time management. Examiners were also less sure about their ability to probe students' thinking and whether the training prepared them adequately.

A total of 179 students (56%) provided additional qualitative feedback on the SCBD in the first implementation year. Of these students, 98 (55%) provided a positive comment related to the SCBD, whereas the remainder provided a negative comment (n = 63, 35%) or a suggestion for improvement (n = 18, 10%) (Table 3). Positive feedback on the SCBD suggested that it was a good assessment format that provided opportunities to demonstrate reasoning and was preferred over other formats such as OSCE. Student concerns about the SCBD related largely to perceived variation between examiners and time management. Students also sought more opportunities for practice and more guidance on the assessment requirements.

Discussion

Our development and implementation of a new standardized oral assessment proved feasible. Acceptability of the SCBD among examiners and students was high, with both groups believing that this new oral assessment allowed exploration of students' DR skills. Examiners used the full range of marks for the SCBD, the assessment had acceptable internal consistency, and the associations with other assessment forms were positive

but not high, suggesting that the SCBD measures a related, yet novel construct.

In this pilot project, we focused on determining the utility of a single case at the end of the third year of the medical course in assessing students' DR. We acknowledge that using a single case limits the format's reliability; however, phased implementation over several years allows for the SCBD to be thoroughly trialled and for a larger bank of cases to be developed for future use. SCBD may eventually comprise several cases in students' assessment. Although developing the SCBD is relatively resource intensive (regarding staff time and the costs of filming and editing the video trigger), this is balanced against the lower cost of administration (eg, there are no costs for simulated patients in this assessment). The lower reliability associated with a single case in the initial implementation is balanced against its high validity as part of an overall assessment programme. We plan to incorporate a similar assessment into our existing OSCE structure at the end of the second year of medical training, where a primary focus is teaching and assessing basic DR.

In constructing the SCBD, undifferentiated cases were developed to allow a very broad range of possible differential diagnoses. This approach maximized different possible avenues for students to explore and afforded students the opportunity to demonstrate their DR. The SCBD emphasized rewarding students for justifying their reasoning more fully than simply making the final diagnosis (which was rewarded if correct).

Examiner variability is a challenge in all forms of performance assessment and students had concerns that the SCBD lacked standardization in this area. The SCBD endeavours to minimize this risk with a comprehensive online training package for examiners and a marking rubric clearly articulating student performance levels. Moreover, an analysis of variation in SCBD scores as a function of examiner suggested that a relatively small percentage of the variation in SCBD scores could be attributed to examiner variation. This suggests reasonable success in our efforts to standardize the assessment. Such findings could be communicated to students to challenge their beliefs that their marks may be overly influenced by individual examiners. To even further improve examiner standardization, we are increasing face-to-face training on exploring students' reasoning and managing transitions between different phases of the discussion. We see value in qualitative research to explore examiner thinking during the assessment, which could include triggered reflection after they watch videos of their own performance. Filming a selection of SCBD performances to undertake cross coding is also an essential next step to affirm the reliability of examiner scoring. We also now have an expanding pool of trained generalist examiners which should also further minimize examiner variability.

In summary, we have developed a highly standardized face-to-face oral assessment allowing in-depth exploration of DR based on an authentic clinical case. It has clear relevance to both examiners and students and will form part of a comprehensive programme of assessment that is designed to drive the teaching and acquisition of important DR skills for our medical students. Such work also has significant implications more broadly in medical education given the continuation of oral examinations as summative assessments in both initial and specialty medical training, in contexts where oral assessments are valued but where there is significant scrutiny to justify the rigour and fairness of assessment processes.

Author's Note

David Smallwood is now affiliated to Austin Hospital, Heidelberg, VIC, Australia and Geoffrey J. McColl is now affiliated to Faculty of Medicine, The University of Queensland, Herston, QLD, Australia.

Author Contributions

RMS, DS, NGC and GJMc devised the innovation and KJR gathered the evaluation data. KJR analysed the assessment and evaluation data and all authors were involved in data interpretation. All authors contributed to the first draft of the manuscript and RMS and KJR revised subsequent drafts in line with author's comments. All authors contributed to revisions of the manuscript and approved the final version for submission.

ORCID iD

Katharine J Reid https://orcid.org/0000-0002-4600-8847

REFERENCES

1. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance test: a tool to assess the reflective clinician. *Teach Learn Med*. 2000;12:189–195.
2. Hrynchak P, Takahashi SG, Nayer M. Key-feature questions for assessment of clinical reasoning: a literature review. *Med Educ*. 2014;48:870–883.
3. Groves M, Scott I, Alexander H. Assessing clinical reasoning: a method to monitor its development in a PBL curriculum. *Med Teach*. 2002;24:507–515.
4. Beullens J, Struyf E, Van Damme B. Do extended matching multiple-choice questions measure clinical reasoning. *Med Educ*. 2005;39:410–417.
5. Joughin G. *A Short Guide to Oral Assessment*. Leeds: Leeds Met Press in Association with University of Wollongong; 2010.
6. Davis MH, Karunathilake I. The place of the oral examination in today's assessment systems. *Med Teach*. 2005;27:294–297.
7. Wass V, van der Vleuten C. The long case. *Med Educ*. 2004;38:1176–1180.
8. Spike N, Jolly B. Are orals worth talking about. *Med Educ*. 2003;37:92–93.
9. Schuwirth LW, van der Vleuten CP. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach*. 2011;33:478–485.
10. Tsukamoto T, Ohira Y, Noda K, Takada T, Ikusaka M. The contribution of the medical history for the diagnosis of simulated cases by medical students. *IJME*. 2012;3:78–82.
11. Pell G, Roberts T. Setting standards for student assessment. *IJRME*. 2006;29:91–103.