Original Article

Development of Computer-based OSCE Re-examination System for Minimizing Inter-examiner Discrepancy

Shinichi Abe* and Eiji Kawada******

** Dental Education Development Center, Tokyo Dental College,
1-2-2 Masago, Mihama-ku, Chiba 261-8502, Japan*

*** Department of Anatomy, Tokyo Dental College,
1-2-2 Masago, Mihama-ku, Chiba 261-8502, Japan*

**** Department of Dental Material Science, Tokyo Dental College,
1-2-2 Masago, Mihama-ku, Chiba 261-8502, Japan*

Received 9 November, 2007/Accepted for publication 26 December, 2007

Abstract

The validity of objective structured clinical examination (OSCE) in evaluating clinical skills has been confirmed in a number of studies. However, inter-examiner reliability has been noted as a concern by many educators. To minimize discrepancy between examiners, in application of the OSCE, we have developed an OSCE attitude and skill test re-examination system. Student performance was digitally, filmed and stored as electronic files. When results showed inter-examiner discrepancy, those items were automatically extracted in the form of a re-examination sheet to allow student re-evaluation by a third examiner in reference to the relevant film files. This system facilitates feedback to each examiner and examinee, offering a potentially useful tool for attitude and skill education.

Key words: OSCE—Re-examination—Inter-examiner discrepancy—Attitude test—Feedback

Introduction

The objective structured clinical examination (OSCE) allows the clinical performance and knowledge of dental students to be evaluated prior to their performing procedures on actual patients^{3,8)}. The validity and reliability of the OSCE in the evaluation of clinical skills have been confirmed^{5,9)}. It is also used to evaluate attitudes⁷⁾ and other types of clinical knowledge in dental students⁶⁾.

The OSCE requires examinees to move through a number of stations and perform

various clinical tasks. In Tokyo Dental College, the examiners is observed and evaluated as they move through each station by 2 examiners. In recent years, however, problems have arisen where examiners have awarded different scores for a particular item⁴⁾. Although, to some extent, problem can be overcome by increasing the number of stations, inter-examiner inconsistency is still an issue. A number of strategies have been proposed to reduce inter-examiner discrepancy: clearer evaluation criteria; avoidance of use of vague expressions in the evaluation manual; record of time

Fig. 1 Schema of system

evaluation made use of a checklist form with subdivided multiple contents; and agreement regarding manual interpretation at the meeting stage prior to carrying out the OSCE. These measures may improve inter-examiner discrepancy. However, since no answer sheet is left with the OSCE, unlike with paper-based tests, there is no accurate re-evaluation method when inter-examiner discrepancy occurs.

To overcome this problem, we have developed a system in which the examinee's performance is digitally filmed, allowing items showing inter-examiner discrepancy to be readily located and re-evaluated.

Methods

1. Outline of system (Fig. 1)

The student's performance is digitally filmed and saved as compressed-image sound files. Each file is registered (year, department, test number, and student' number) on a database server. The system automatically extracts cases where the students does not obtain a prerequisite mark, or where an evaluation discrepancy is detected between examiners for a given item.

The relevant file and score are displayed on

a PC screen which may be accessed via the database by year (Christian Era), department, test number(that is, whether this is the first, second, or third time, etc., for a given item; hereafter, TN), and student number. The re-examiner performs a re-evaluation using a re-examination sheet and determines the final score.

2. Specifications

1) Data file

The question files used are "Comma Separated Value (CSV) files produced by Microsoft Excel. These may be cataloged solely by year (Christian Era), department, and TN.

Unique test result data are cataloged by year (Christian Era), department, TN, and student' number.

2) Data file storage

The files, test results, and questions are stored in the designated locations on a database server (Video On Demand (VOD) server).

3) Access to data file

Access to the re-examination sheet program is restricted by a "username" and "password" which are determined in advance.

4) Access to rating data

Rating result data are accessed in a data file by an arbitrarily attributed file name. This file

Fig. 2 Example of re-examination sheet

On re-examination sheet, relevant film data are accessed by clicking "Image start" button.

is an Excel file (Microsoft), and can be printed out using Excel.

5) Production of question files

The production platform used is Excel (Microsoft). The files are "CSV" files.

6) Test sheet

The required question file is selected from the list of question files according to department. It is then downloaded from the database server, and displayed.

7) Initiation and discontinuation of test

The test is initiated using a "Test start" button. Unique data on each test (year, department, TN, and student' number) are displayed and confirmed.

The test is terminated by clicking the "Test start" button again. However, when no score has been given, a warning is displayed.

8) Recording/discontinuation of recording

To record, the "During recording" button is clicked. The initiation of recording is not simultaneous with the initiation of the test. However, recording cannot be initiated until the unique data for a given test have been confirmed.

Recording is discontinued by clicking the "Recording stop" button. The recording is

always registered as VOD on the database server. However, when no score has been given, a warning is displayed.

9) Scoring

Clicking the "Score" button allows input of the score for the relevant question file. Whenever the "Score" button is clicked, the "test results" content score subtotals, and the total score is displayed in the lower area of the screen. Scores can be freely changed. When a score is given, the results are transferred to the database server. A score can be changed as many times as necessary. When a question is not scored, a warning is displayed.

10) Re-examination sheet (Fig. 2)

A list of result data files is displayed based on year (Christian Era), department, TN, and the necessary file is selected. There are 4 conditions for listing file data, as follows: (1) students who failed the test; (2) cases where an inter-examiner difference in the score according to the criterion mentioned above occurred; (3) those who passed the test; and (4) all students. Films are accessed by clicking the "Image start" button, and may be viewed in standard reproduction mode, pause, forward, fast-forward, rewind, fast-rewind, or

Table 1 Medical interview sheet

	2	1	0
1. Interview process			
1) Can give greetings	☐		☐
2) Informing the patient of what is going to be performed	☐		☐
3) Explaining to the patient that it is okay to freely ask questions	☐		☐
4) Appropriate body language	☐		☐
5) Confirmation of the degree of the guardian's understanding	☐		☐
6) Use of simple words	☐		☐
7) Way of talking that is easy too understand	☐		☐
8) Confirmation of questions	☐		☐
2. Interview contents			
1) Can explain the effects of fluorine	☐	☐	☐
2) Can explain the treatment technique	☐	☐	☐
3) Can explain cautionary items after treatment	☐	☐	☐
Total score:			

change in stop/reproduction site by movement of a slide bar.

11) Re-examination

Clicking the “Re-examination” button allows viewing of the re-examination score given for the relevant question file. The added score and normalized score (%) are displayed each time.

Results

1. System performance

The system worked without problems during the OSCE. Examiner evaluation and the relevant image file were linked without any problem. After the test, students showing an inter-examiner difference of 10% or more in the score were automatically extracted, and re-examination was performed. Re-examination was performed without difficulty.

2. OSCE practice at Department of Pediatric Dentistry

Instructors were required to role-play the mothers of minors undergoing a medical interview. The contents of the interview and subsequent evaluation are shown in Table 1. Forty-seven students were divided into 2 groups (23 and 24 students). At each station, their

Fig. 3 OSCE practice

Evaluation of medical interview with contents shown in Table 1 was performed.

performance was evaluated by two examiners. Before the evaluation, 4 raters practiced for about 30 minutes to accustom themselves to the system (Fig. 3).

Discussion

As of fiscal year 2006, the OSCE was adopted as the common achievement test to be performed under uniform conditions throughout all the 28 dental colleges in Japan. This means that all colleges use the same evalua-

tion sheets, evaluation manual, and management/practice guidelines. These evaluations were already being performed for some years prior to this, and a number of problems were highlighted and discussed. At present, neither inter-examiner consensus nor re-evaluation after examination are possible with this system. However, inter-examiner discrepancies in scoring for items related to attitude have presented problems and have been discussed⁹⁾. Cohen's Kappa value¹⁾ has been used as a parameter to determine inter-examiner agreement. This value is calculated by taking into consideration the probability of incidental agreement with the percentage of agreement among multiple examiners. Ito *et al.*⁴⁾ reported that the Kappa value sometimes increases, or its calculation becomes impossible even when the rate of agreement is the same. They suggested the importance of overcoming these problems and coming to a consensus on the description of methods in the evaluation manual and formulating definite evaluation criteria.

However, previous studies discussing inter-examiner agreement^{2,7)} have given no consideration to re-examining when the rate of agreement is low. This was because re-evaluation of the test was previously impossible with the OSCE, unlike with paper tests. We considered this disadvantage to be the most important problem with OSCE when developing our system, and used a third examiner as a "re-examiner" when the rate of inter-examiner agreement was low. In addition, our aim was to develop a system that facilitated re-examination.

In the specifications of this system, special consideration was given to the following points.

Question files produced using Excel (Microsoft) before carrying out the OSCE were stored as data files (CSV files). This facilitates input of evaluation items in each test, because the contents of each evaluation item differ among departments. Since question files are stored in the data server, the necessary file can be easily accessed without confusion by inputting the year (Christian Era), department, and TN.

The test sheet is selected from the list of question files, and the necessary question file is downloaded from the database server and displayed. Examiners input the relevant student number. We believed that confusion was likely to occur during totaling if examiners inputted the wrong number. Therefore, this system was designed to allow both input and selection from a drop-down list. Before the test, each examiner can set the range of this list to include only those students with the examiner is concerned. This facilitates selection, as the examiner only has a small range to choose from.

To total results, not only the total of raw scores, but also the score converted on a maximum scale of 100 can be automatically displayed as a list. Since weighting may differ among questions, the system allows each question to be awarded an appropriately weighted score at the time of inputting questions.

Students who failed the test according to the minimum passing score arbitrarily determined by each department, and students showing an inter-examiner difference above the permissible degree of discrepancy, are automatically extracted and displayed on the re-examination sheet on the PC screen. Here, after confirmation of relevant film and score by the two initial examiners, another examiner performs re-evaluation and inputs the final score as the re-examination score. The possibility of re-examination to achieve an accurate score allowed the problems inherent in the standard OSCE to be overcome.

This system was used in an actual OSCE in the Department of Pediatric Dentistry. Items that produced inter-examiner discrepancies were extracted. By using film, it was possible to obtain feedback on evaluation, which may prove useful in determining evaluation criteria and educating examiners in the future. In addition, this system allowed the students themselves to access feedback on their performance using the re-examination sheets. Students can observe themselves or others undergo the OSCE, and narrow that access down to specific problem areas, where required, which offers a potentially useful

tool in student education.

In this study, we only reported the OSCE re-evaluation system itself. A detailed study on its effectiveness and results of its implementation will follow in a further report.

Conclusions

To overcome inter-examiner discrepancies in OSCE attitude evaluation, we developed the following OSCE performance test re-examination system.

- 1) The OSCE performance test is digitally filmed in real time and stored as electronic files.
- 2) Evaluation items are arbitrarily inputted into Excel prior to application of OSCE; they are then organized and stored as question files.
- 3) Items showing inter-examiner disagreement are automatically extracted as a re-examination sheet; a re-examiner confirms the scores of the first two examiners based on the relevant film data, performs a re-evaluation, and can correct the earlier score, where necessary.
- 4) A re-evaluation sheet allows feedback to both examiners and students.

Acknowledgements

Part of this study was supported by a Grant-in-Aid for University Reform under the Support Program for Distinctive University Education in 2005 from the Ministry of Education, Science, Sports and Culture. We would like to express our deep gratitude to Mr. Nobuyuki Sahori, Sony Broadband Solutions Corp. and Mr. Kuniaki Nose, Com Science Corp. for their cooperation in the development of this system. We would also like to thank the members of the Department of Pediatric Dentistry, Tokyo Dental College.

References

- 1) Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:37–44.
- 2) Feather A, Kopelman PG (1997) A practical approach to running an Objective Structured Clinical Examination (OSCE) for medical undergraduates. *Educ for Health* 10:333–350.
- 3) Harden RM, Steveson M, Downie WW, Wilson GM (1975) Assessment of competence using objective structured examination. *Br Med J* 1:447–451.
- 4) Ito T, Aoki S, Oomine H, Kawai Y, Kasai K, Kanazawa E, Ootake S (2005) OSCE evaluation discrepancies in medical interview settings —Based on an analysis of human errors—. *J Jpn Dent Educ Assoc* 21:21–30.
- 5) Kakehashi Y, Ogiso B, Nakajima I, Gionhaku N, Kuwata F, Otsuka K (2005) Coincidence and effect of feedback on evaluation of three tests used for objective structured clinical examination (OSCE) in prosthodontics. *J Jpn Dent Educ Assoc* 21:10–20.
- 6) O'Connor HM, McGraw RC (1997) Clinical skills training: developing objective assessment instruments. *Med Educ* 31:359–363.
- 7) Prislin MD, Fitzpatrick CF, Lie D, Giglio M, Radecki S, Lewis F (1998) Use of an objective structured clinical examination in evaluating student performance. *Fam Med* 30:338–344.
- 8) Rosebraugh CJ, Speer AJ, Solomon DJ (1997) Setting standards and defining quality of performance in the validation of a standardized-patient examination format. *Acad Med* 72: 1012–1014.
- 9) Suganuma T, Kakuta S, Hasegawa T, Kimura Y, Miyashita H, Amano H, Tsukasaki H, Gotoh K, Okano T, Sano T (2005) Evaluation agreement in objective structured clinical examination. *J Jpn Dent Educ Assoc* 21:3–9.

Reprint requests to:

Dr. Shinichi Abe
Department of Anatomy,
Tokyo Dental College,
1-2-2 Masago, Mihama-ku,
Chiba 261-8502, Japan