

Variation of Scores Among Examination Centers in Short Case Examination

Nitipatana Chierakul, M.D.*,**, Paka Kontee, B.BA*, Chana Naruman, M.Sc.**

*Subcommittee for Training and Examination, The Royal College of Physicians of Thailand, **Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand.

ABSTRACT

Objective: To evaluate the impact of patient selection process by examination center and examiner subjectivity on scores from the Royal College of Physicians Thailand (RCPT) short case examination.

Methods: Data of candidates who encountered short case examination as one part of clinical examination in RCPT board certifying examination in academic year 2010 were retrieved. For each examination center, overall scores given by internal and external examiners were compared. Among patient categories that were selected for the examination in all examination centers, comparison between scores in each category were also explored.

Results: There were 25 examination centers involved with 24 patient categories encountered. The mean score from internal and external examiners ranged from 64.0 ± 17.1 to 77.8 ± 14.4 and 65.6 ± 18.9 to 77.4 ± 15.4 respectively. Stringent site categorization as determined by lower scores from internal examiners were established in 11 examination centers, and, vice versa, 15 were categorized as lenient site. The lowest mean score difference in stringent sites was -4.9 and the highest from lenient sites was +4.7, but no statistical significances were reached ($p = 0.088$ and 0.106 respectively). There were 2 patient categories which yielded a significantly higher score and 1 patient category with lower score ($p < 0.05$).

Conclusion: In the current RCPT short case examination, the selected patient category appears to contribute variation in scores more than examiner subjectivity.

Keywords: Short case, patient selection process, examiner subjectivity

Siriraj Med J 2014;66:53-55

E-journal: <http://www.sirirajmedj.com>

INTRODUCTION

The Royal College of Physicians of Thailand (RCPT) board certifying examination consists of written and clinical parts. Minimum passing level is used to determine the success, if the trainees partially fail, they can reexamine only the failed part within 3 academic years. Long case, short case, and laboratory interpretation have been nominated as tools to assess clinical competency in this high-stakes examination.

The authors have previously demonstrated modest relationship between the scores from this clinical examination and the scores from written examination.¹ Acceptable inter-rater reliability among a pair of short case examiners in this final board certifying examination was

also established.² Short case has been traditionally used as a tool of clinical assessment both in undergraduate and postgraduate medical education. At present, short case is usually transformed into one part of the objective structured clinical examination (OSCE). In this study, we aim to evaluate the effect of examination center on scores from short case examination.

MATERIALS AND METHODS

The RCPT short case examination

There are 2 occasions of RCPT clinical examination during the third year of internal medicine residency training, and each candidate must perform 3 encounters of short case examination on each occasion which contribute 27% of the total score for clinical examination. During an occasion, each examination center independently chooses 2 physical examination stations in the field of cardiology or neurology or dermatology. The other station may be a physical examination of patient in another field or be tested by giving the health information to a simulated

Correspondence to: Nitipatana Chierakul
E-mail: nitipat7@gmail.com

Received 2 July 2013

Revised 19 September 2013

Accepted 24 September 2013

patient. For each station, an individual examination center can also independently select 1 from 6 patient categories that has been specifically assigned by the RCPT for each occasion. In each examination center, the RCPT randomly allocates 6 or 12 candidates from various training centers for performing the encounter. Every candidate in each examination center must perform the task with the same patient categories that may be similar or different from other examination centers.

In each station there were 2 examiners, one from the examination center and another from a different hospital which has an internal medicine training program. The examiners are appointed by the RCPT, and they must have age between 35-65 years, have internal medicine or related board certification, and have worked as an internist for more than 10 years (including years in training). Instruction for the expected role and structured marking form is distributed to each examiner by the representative of each training center at least 4 weeks before the examination.

During an encounter, a candidate is directly observed and interacts with the examiners for a period of 8 minutes. Both internal and external examiners independently make notes on a 5-level standard structured marking scale prepared by the RCPT for both of the process and the outcome of task performed by the candidate. Before the assessment, both examiners spent 10 minutes with the patient to confirm or adjust the clinical note prepared by examination center.

Data collection

Data from internal medicine candidates who performed the clinical part of RCPT board certifying examination in academic year 2010 were collected. For each examination center, mean scores from internal and external examiners were compared using *t*-test. Comparison of mean scores in each patient category that was selected was performed using 1-way ANOVA and Tamhane method for post-hoc multiple comparisons. All statistical analyses were carried out by statistical software SPSS version 13.0 (SPSS Inc, Chicago, USA).

RESULTS

In that academic year, RCPT appointed 25 examination centers for the clinical examination. There were 24 patient categories selected for 1,548 encounters on two occasions. The range of scores (mean±SD) given overall by internal and external examiners were 64.0±17.1 to 77.8±14.4 and 65.6±18.9 to 77.4±15.4 respectively, with no statistical difference (*p* = 0.863). We considered the examination center with lower scores from internal examiners as a stringent site (Table 1), the lowest mean score difference in this group was -4.9. However scores from the most stringent ones had no statistical difference from the rest. On the contrary, a lenient site with higher scores from internal examiners is shown in Table 2. Internal examiners in this group provided more scores up to +4.7, but the statistical significance for the difference was also not reached.

TABLE 1. Mean and standard deviation of score from internal and external examiners among stringent examination centers.

Center	Internal score	External score	Difference	p-value
1	64.0±17.1	68.9±17.2	-4.9	0.088
2	71.3±16.1	75.4±17.6	-4.1	0.149
3	68.4±18.2	71.8±17.5	-3.4	0.421
4	70.9±15.6	73.9±14.1	-3.0	0.219
5	71.9±16.0	74.9±17.4	-3.0	0.279
6	73.4±12.1	76.0±12.1	-2.6	0.370
7	72.0±13.6	73.6±16.6	-1.6	0.561
8	72.1±15.8	72.9±16.3	-0.8	0.742
9	69.6±18.9	70.4±20.3	-0.8	0.796
10	76.8±14.7	77.4±15.4	-0.6	0.764
11	67.8±16.5	67.9±18.1	-0.1	0.990

Among 24 patient categories that had been selected for the short case, scores from two patient categories were significantly higher than the other 11 categories, and from only one with significant lower scores than the other 13 categories as in Table 3 (*p* < 0.05).

DISCUSSION

Current RCPT short case in the OSCE format aims to assess performance which resembles actual doing in clinical practice. Although the candidate has a chance to perform the examination on two occasions at two different examination centers, for dispersion of the random chance by “luck of the draw”, the effect of patient selection among different examination centers does exist. Future methods for score adjustment from “Hawk” and “Dove” examination centers will provide more fairness, especially for those marginal candidates. Reduction of the chance to choose which patient category for an encounter by an individual examination center, may also be an option. For maturation effect of the candidates from different occasions at 6 months apart as demonstrated in the authors previous study², this positive gain was equally distributed and should not impact the result of the study.

TABLE 2. Mean and standard deviation of score from internal and external examiners among lenient examination centers.

Center	Internal score	External score	Difference	p-value
12	66.0±16.2	65.9±19.3	+0.1	0.976
13	69.3±17.4	69.2±17.5	+0.1	0.970
14	74.5±12.7	74.3±12.4	+0.2	0.955
15	69.7±19.1	69.5±19.7	+0.2	0.954
16	74.2±16.6	73.7±17.1	+0.5	0.896
17	71.8±17.4	70.4±17.2	+1.4	0.735
18	77.8±14.4	75.9±13.0	+1.9	0.568
19	76.0±16.1	73.6±15.5	+2.4	0.364
20	73.6±16.4	70.9±15.4	+2.7	0.296
21	68.6±22.5	65.8±18.1	+2.8	0.413
22	76.9±17.0	72.9±14.8	+4.0	0.214
23	77.3±17.6	73.2±21.5	+4.1	0.387
24	69.9±17.5	65.6±18.9	+4.3	0.158
25	73.1±15.3	68.4±15.0	+4.7	0.106

TABLE 3. Mean and standard deviation of score from selected patient categories.

Category	No. of encounter	Score
1	96	71.5±14.0
2	120	71.9±17.9
3	150	70.7±18.0
4	30	65.0±14.1
5	12	61.6±14.3
6	18±	61.0±14.9
7	96	64.7±17.4
8	51	64.0±14.5
9	60	72.4±16.0
10	45	75.0±14.4
11	60	75.2±13.6
12	6	62.0±8.3
13	36	66.4±18.7
14	52	81.7±12.9 *
15	24	82.3±9.9 *
16	24	69.1±19.9
17	56	63.5±18.4
18	126	77.8±15.4
19	90	65.0±18.2
20	66	78.3±12.7
21	18	75.9±11.5
22	12	51.0±10.9 *
23	90	73.3±12.9
24	210	74.0±12.1

*p <0.05

In terms of patient heterogeneity, the OSCE format of short case can reduce this bias by using direct observation and a structured manner for scoring.³⁻⁶ However, the opportunity for the examination center to

select patient category, can create both risk or benefit to the candidates. Adjustment maneuver and reevaluation of the properness of those “Black sheep” and “Golden sheep” patient categories must also be performed. A single nationwide OSCE on one occasion can reduce this bias. However, this is labor intensive, and lessons learnt from the organization of OSCE in the National Medical License Examination of Thailand can be applied for their feasibility.

CONCLUSION

Patient selection process by the examination center contributes greater variation in scores than the examiner subjectivity in current RCPT short case format as part of a board certifying examination.

REFERENCES

1. Chierakul N, Danchaivijitr S, Kontee P, Naruman C. Relationship between outcome of written and clinical parts in internal medicine board certifying examination. *Siriraj Med J*. 2009 Jul;61(4):194-6.
2. Chierakul N, Danchaivijitr S, Kontee P, Naruman S. Reliability and validity of long case and short case in internal medicine board certification examination. *J Med Assoc Thai*. 2010 Apr;93(4):424-8.
3. Holmboe ES, Hawkins RE. Methods for evaluation the clinical competence of residents in internal medicine: a review. *Ann Intern Med*. 1998 Jul 1;129(1):42-8.
4. Holmboe ES, Hawkins RE, Huot SJ. Effect of training in direct observation of medical residents' clinical competence. *Ann Intern Med*. 2004 Jun 1;140(11):874-81.
5. Wilkinson TJ, D'Orsogna LJ, Nair BR, Judd SJ, Frampton CM. The reliability of long cases and short cases undertaken as practice for a summative examination. *Intern Med J*. 2010 Aug;40(8):581-6.
6. Boursicot K, Etheridge L, Setna Z, Sturrock A, Ker J, Smee S, et al. Performance in assessment: Consensus statement and recommendations from the Ottawa conference. *Med Teach*. 2011 May;33(5):370-83.