Available online at www.sciencedirect.com

ScienceDirect

Biomedical Journal

journal homepage: www.elsevier.com/locate/bj

Original Article

Comparing the outcomes of different postgraduate year training programs in Taiwan

Peng-Wei Hsu [a], *Ming-Ju Hsieh* [b,c], *Ren-Huei Fu* [b,d], *Jing-Long Huang* [d], *Mei-Chen Liao* [b], *Shih-Tseng Lee* [a,*]

[a] *Department of Neurosurgery, Chang Gung Memorial Hospital at Linkou, Chang Gung University College of Medicine, Taoyuan, Taiwan*
[b] *Department of Medical Education, Chang Gung Memorial Hospital at Linkou, Chang Gung University College of Medicine, Taoyuan, Taiwan*
[c] *Department of Chest Surgery, Chang Gung Memorial Hospital at Linkou, Chang Gung University College of Medicine, Taoyuan, Taiwan*
[d] *Department of Pediatrics, Chang Gung Memorial Hospital at Linkou, Chang Gung University College of Medicine, Taoyuan, Taiwan*

ARTICLE INFO

Article history:
Received 30 July 2014
Accepted 31 July 2015
Available online 9 March 2016

Keywords:
Checklist
Multiple choice question
Objective structured clinical examination
Postgraduate training
Standardized patient

ABSTRACT

Background: Postgraduate year training programs play an important role in the development of a comprehensive medical education. The goal of these training programs is to inculcate in physicians the expected level of skill in patient care. After the initiation of such programs in the USA, Europe, and Japan, studies were conducted in Taiwan to investigate relevant training methods, and a training system was established in 2003. Beginning with 3-month programs, followed by 6-month programs, the programs were constantly modified and enhanced by the establishment of the 1-year training program in 2011. This year was the transition period from the 6-month programs to the 1-year programs.
Methods: We used a 50-item multiple choice question (MCQ) test and six 10-min stations for objective structured clinical examination (OSCE), which was composed of four stations relating to standardized patients and two stations concerning the clinical skill evaluation, to evaluate the learning results of the trainees. The trainees were divided into four groups according to the training program.
Results: There was no significant difference between the performance of the 6 months and 1-year groups. The p values were 0.424 in the MCQ test and 0.082 in the OSCE evaluation.
Conclusion: A well-designed postgraduate training program should develop trainees' competencies. The results of this study may provide useful insight for ways to improve the design of training programs. Further investigation to better understand the impact of different programs is warranted.

* *Corresponding author*. Department of Neurosurgery, Chang Gung Memorial Hospital at Linkou, 5, Fusing St., Gueishan, Taoyuan, Taiwan, ROC. Tel.: +886 3 3281200x2119; fax: +886 3 3285818.
E-mail address: ns3096@adm.cgmh.org.tw (S-T. Lee).
Peer review under responsibility of Chang Gung University.
http://dx.doi.org/10.1016/j.bj.2016.01.006

At a glance commentary

Scientific background on the subject

PGY training programs play an important role in the development of a comprehensive medical education. The goal of these training programs is to inculcate in physicians the expected level of skill in patient care. The training programs were conducted in Taiwan to investigate relevant training methods, and a training system was established in 2003. Beginning with 3-month programs, followed by 6-month programs, the programs were constantly modified and enhanced until the establishment of the 1-year training program in 2011. This year was the transition period from the 6-month programs to the 1-year programs.

What this study adds to the field

This study we performed revealed no significant differences in medical knowledge and clinical performance among the trainees regardless of different programs, and showed that the learning results persisted long after the training ended if the programs were well designed. The results provide valuable information that can be used to improve the design of the training program such as arranging more core competencies in the 1-year program.

In the past, medical students were trained through knowledge-centered learning to obtain clinical skills and develop patient care abilities. The lack of general medical training was accepted due to the immediate entry into a specialization after graduating from school [1]. Over time, more advanced countries set up 1–2 years general medical training programs in order to improve the abilities of patient care, clinical skills, doctor-patient communication, and the integration of the patient care experience. The postgraduate general practice training system was set up as a pre-registration year in the United Kingdom in 1951 [2]. The postgraduate year (PGY) training program developed in America in 1970 [3]. In Taiwan, the doctor training program began in 1897 and from 1950, followed the American format. Medical students graduated from medical school after completing 6 years of college education and 1-year of internship training and then chose their specialty in the residency training program. The PGY training program was launched in Taiwan in 2003, starting with a 3-month course that eventually progressed into a 1-year course in 2011.

However, little is known about the effectiveness of this program in Taiwan on trainee learning. We also want to evaluate the impacts of the different training programs and timing issue on the learning effect. The objective of this study, therefore, was to use various assessment tools to compare clinical core competencies and relative attitudes to postgraduate general medicine practice among PGY residents at Chang Gung Memorial Hospital. We hope that the quantitative data and the qualitative information can be used to improve program design and accurately evaluate the implementation of postgraduate general medicine training in Taiwan.

Materials and methods

Participants were 314 trainees. The evaluation was conducted in two parts. The first consisted of a 50-item multiple choice question (MCQ) test with each item worth two points. The items were chosen according to the core knowledge required, as indicated by the Taiwan Joint Commission on Hospital Accreditation (TJCHA). The second part involved six 10-min stations objective structured clinical examination (OSCE). Standardized patients (SPs) were used in four stations including internal medicine, surgery, obstetrics and gynecology, and pediatrics. Two stations concerned clinical skills performance such as endotracheal tube intubation and infection-protective clothing. The evaluation was held in the last month of the training program the Group PGY trainees accepted mentioned as below.

All of the 314 trainees participated in the MCQ exam. They were divided into four groups according to their training program.

Group R2 contained 156 2nd-year residents enrolled in a 6-month PGY training program.

Groups R1a and R1b contained 61 and 49 1st-year residents, respectively who were also enrolled in a 6-month PGY training program. According to the TJCHA's policy, the 61 R1a residents were enrolled in the PGY training program from July to December 2011 and then continued onto their 1st-year resident training program. The 49 R1b residents proceeded with their resident training program and then enrolled in the PGY training program from January to June 2012.

Group PGY consisted of 48 general residents who had just completed their internship training and then enrolled in a 1-year PGY training program from July 2011 to June 2012.

In Groups R2, R1a, and R1b, the residents chose their specialization for residency prior to enrolling in the PGY training program. The trainees of Group PGY had not decided on their specialization for a residency at the time the study was conducted.

In the second part, 24 residents from each group ($n = 96$) chosen randomized participated in the OSCE. The criteria for passing or failing each station were determined by the Angoff method. The results of every checklist were divided into three possible scores, not completed (score of 0), partially completed (score of 1), and fully completed (score of 2). The final score obtained at each station was determined by using the following equation: (Score obtained/maximum obtainable score) $\times$ 100. The mean score was then calculated across all stations. All the raters were qualified by the Taiwan Association of Medical Education after completing the rater training program.

The item difficulty index and the item discrimination index of the MCQ test were analyzed after the assessment. The trainees were scored by arrangement, taking the upper and lower quartiles, and then categorized into high- and low-grade groups with respect to the correct rate for each item as percentage in high (PH) or percentage in low (PL). The item difficulty index was calculated as (PH + PL)/2 and the item

discrimination index as (PH−PL). An unpaired t-test, ANCOVA, and Pearson correlations were used to analyze the data via SPSS Version 19.0 (SPSS, Inc., Chicago, IL, USA). A p value below 0.05 was considered to indicate statistical significance.

Results

The mean MCQ score for all of the 314 doctors was 68 ± 7 (range: 40–86). After further analysis, the mean scores in the four groups were 68 ± 7 (range: 52–82) in the R2 group, 69 ± 7 (range: 40–86) in the R1a group, 68 ± 8 (range: 48–86) in the R1b group, and 69 ± 7 (range: 46–86) in the PGY group. There was no significant difference between the four groups ($p = 0.424$).

The passing rates of the first and last 25% were used to determine the item discrimination and difficulty index for the MCQ test. The item discrimination index was defined as follows: Bad (≤0.19), acceptable (0.2–0.29), good (0.3–0.39), and excellent (≥0.4). The item difficulty index was defined as difficult (<0.4), moderate (0.4–0.6), and easy (>0.6). Among the 50 MCQ items, the item discrimination index was bad in 27 (54%), acceptable in 11 (22%), good in six (12%), and excellent in six [12%; Fig. 1A]. We re-evaluated the trainees' performance after excluding the 27 items with a bad index. The mean number of items passed was 13.2 in Group R2, 13.8 in Group R1a, 12.8 in Group R1b, and 13.5 in Group PGY. There was also no significant difference [$p = 0.429$, Fig. 2]. The item difficulty index was easy in 9 (39%), moderate in 10 (44%), and difficult in four (17%) of 23 items [Fig. 1B].

In the OSCE, the mean final scores of the six stations were 64.6 ± 6.5 in Group R2, 64.9 ± 6 in Group R1a, 64.1 ± 6.2 in Group R1b, and 68 ± 4.8 in Group PGY. The p value was 0.082 for the four groups [ANCOVA, Fig. 3]. When the performance difference between the assessments was analyzed, the p values were 0.236 for the SP-stations assessment and 0.527 for the clinical skills performance assessment. Finally, the correlation coefficient between the MCQ and the OSCE of all trainees was 0.333 [$p = 0.002$, Fig. 4].

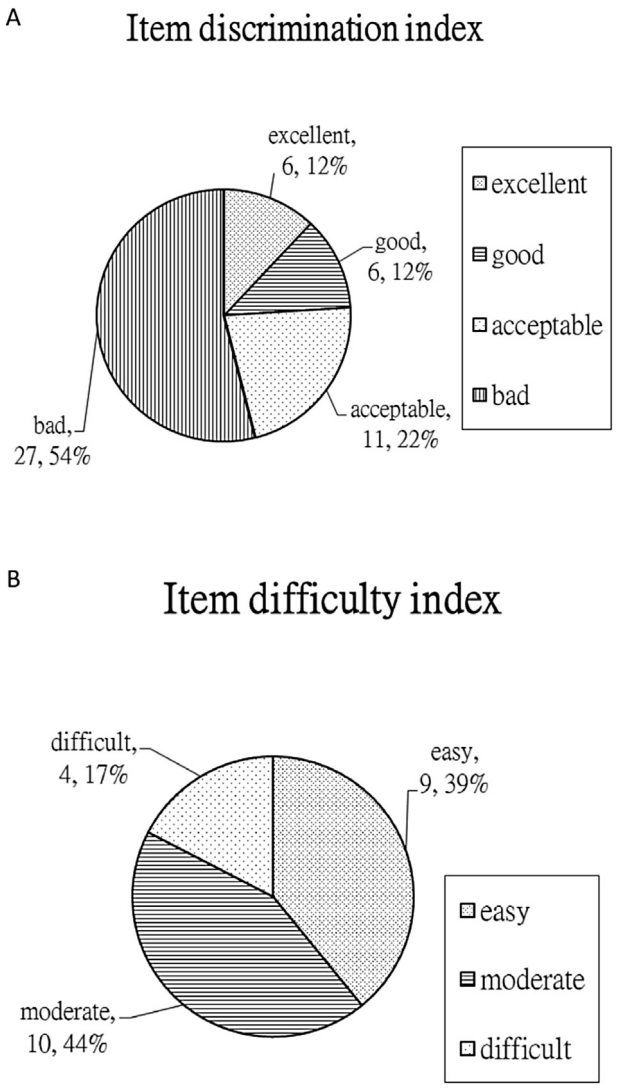

Fig. 1 – (A) Evaluation of the discrimination index for the 50 multiple choice question items. (B) After items with a bad discrimination index were excluded, the difficulty index was calculated for 23 items.

Discussion

The postgraduate training program for general medicine was implemented by the Taiwanese government after the severe

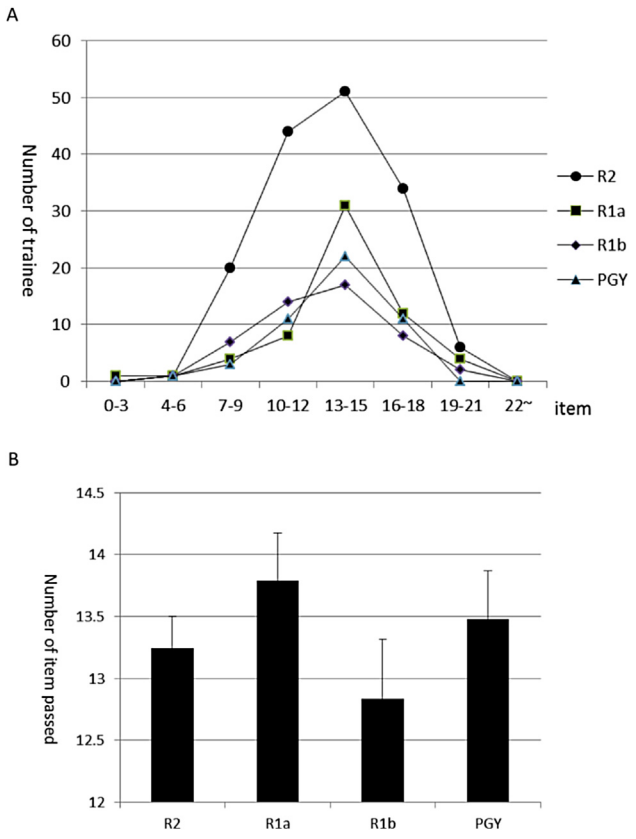

Fig. 2 – (A) Trainee distribution for the 23 multiple choice question items with excellent, good, and acceptable discrimination indices. (B) There were no significant differences among groups with regard to the number of items passed (p = 0.429).

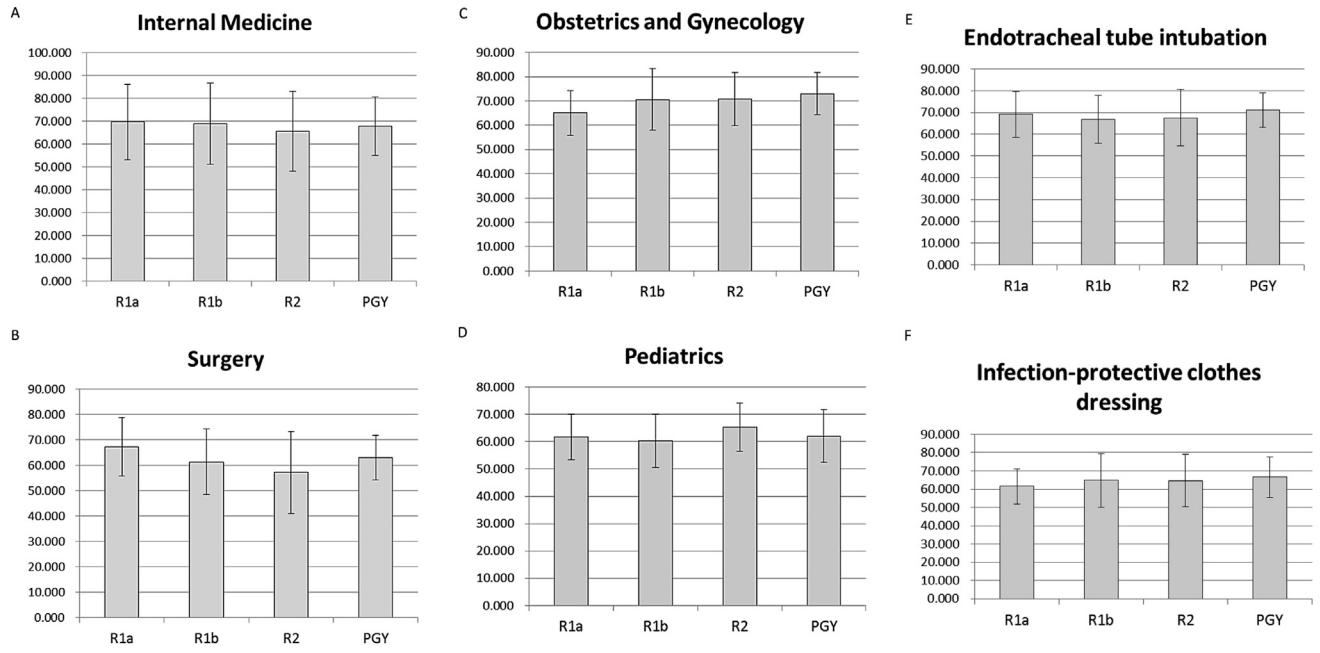

Fig. 3 — The results and comparisons of the objective structured clinical examination among the four groups. (A) Internal medicine ($p = 0.834$); (B) surgery ($p = 0.297$); (C) obstetrics and gynecology ($p = 0.071$); (D) pediatrics ($p = 0.633$); (E) endotracheal tube intubation ($p = 0.525$); and (F) infection-protective clothing ($p = 0.575$). No significant difference was found among groups.

acute respiratory syndrome pandemic in 2003 to address a need for improved professional training. The training program was implemented by the TJCHA with the aim of improving the competency of medical graduates with respect to patient-centered care as well as developing their the ability to perform holistic medical care and competency in medical knowledge, clinical skills, professional attitude, etc. PGY residents in Taiwan have been required to complete a general medicine training program since August 2003.

The current form of the Taiwanese postgraduate training program developed over three stages. The initial stage of the training program included a 3-month training period where the goal was to improve medical graduates' knowledge and attitude toward community health. After July 2006, the PGY program was extended to incorporate a 6-month training course (the second stage). It included the development of the training model and assessment methods and consisted of 1-month of training in general medicine, 2 months of training in community medicine, 3 months of training in specialty courses focused on primary care, and was followed by another 6 months of training in holistic care practice. In the third stage, a full-year program and was initiated in August 2011. This program included 3 months of community medicine, 3 months of general medicine, 2 months of general surgery, 1-month of emergency medicine, 1-month of pediatric medicine, 1-month of obstetrics and gynecology, and 1-month of a chosen specialty course.

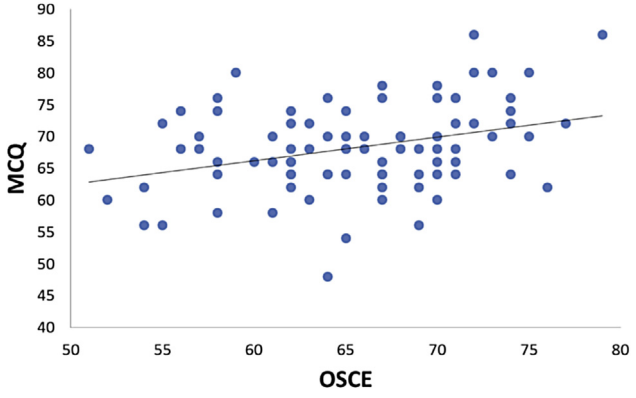

Fig. 4 — The correlation coefficient between the multiple choice question and the objective structured clinical examination of all trainees was 0.333 ($p = 0.002$).

In the first and second stages, the students could choose a specialized residency after graduating from medical school with the PGY training program being included in the 1st year of the residency training program. In the third stage, the students became general medical residents after graduation and enrolled in the full-year PGY training course prior to choosing a specialty residency. In 2011, there was an overlap of the second and third stages of the PGY training program, which provided a good opportunity to analyze and compare the results of the two programs.

The six core competencies emphasized and cultivated in the PGY training program followed the rules suggested by the Accreditation Council for Graduate Medical Education (ACGME). These competencies were patient care, medical knowledge, professionalism, interpersonal and communication skills, practice-based learning and improvement, and systems-based practice. It was important that the program had an effective plan for assessing trainees' performance throughout the program and a method for utilizing

assessment results to improve the residents' performance. An evaluation toolbox from the ACGME suggested the best methods to assess competence [4]. SPs, checklists, and OSCEs were used to evaluate competency in interpersonal and communication skills and patient care. MCQs and oral examinations were useful for evaluating competency in medical knowledge while OSCEs and checklists evaluated professionalism effectively. Student competency in practice-based learning and improvement was assessed with OSCE, SPs, checklists, and MCQ tests. In the evaluation of system-based practice, MCQ, OSCE, and checklists proved useful. Given their proven efficacy, MCQ tests, OSCEs with SPs, and checklists were used to analyze the learning outcomes of the different training programs examined in this study.

The MCQ test was used in departmental or comprehensive examinations for determining progress or certification [5]. It is used more widely than the other methods due to its cost-effectiveness and its ability to yield a reliable score. The effectiveness of the MCQ test depends on a close relationship between the quality of the overall examination and the individual items. Each item should be developed to test competence in a clinical situation or in handling laboratory data, not memory, and students should be required to apply the knowledge they have gained to find a solution to the problem presented. Guidelines on the development of such items have been published [6,7]. Furthermore, the structure of the items plays an important role in their discriminatory power. Jozefowicz et al. presented a scale for rating item quality [8]. All of the items, we used were developed following the aforementioned principles and had matched at least score four on Jozefowicz et al.'s scale. One of the most important aspects of quality is the ability to discriminate between students who learn well and those who did not. The discrimination index is also a valid measure of item quality [9]. A relationship has also been demonstrated between the item discrimination index and the difficulty index [10]. We analyzed the discrimination index of the original 50 items, with 23 items (46%) having acceptable results, and nearly half of the 23 items being moderately difficult (10 items, 44%). Though, there was no difference between the trainees after the evaluation via the 23 items, more items were needed to confirm the result.

A useful assessment tool is the use of SPs in a simulated clinical encounter, otherwise known as the OSCE. The OSCE was first introduced by Harden and Gleeson in 1979 [11]. Interactions with SPs can be tailored to meet specific educational goals and student performance can be rated dependably [12]. According to the literature, evaluation reliability could be increased from 0.85 to 0.90 if there is a sufficient number of stations and trainees [13]. The specific skills rated during the OSCE at our institute include history taking skills, physical examination skills, communication skills, technical skills, and skills on data interpretation, differential diagnosis, and making treatment decisions. Through the use of a checklist after evaluation of its reliability and validity, it could provide an objective and organizational structured assessment of trainees' technical skills [14]. Across the modalities, there were no statistically significant differences among the four groups in our study and the trainees performed similarly after a further analysis (the p values were 0.834 for internal medicine, 0.297 for surgery, 0.071 for obstetrics and gynecology, 0.633 for pediatrics, 0.525 for endotracheal tube intubation, and 0.575 for infection-protective clothing).

Conclusion

This study revealed no significant differences in medical knowledge and clinical performance among the four groups of trainees regardless of program, and showed that the learning results persisted long after the training ended if the programs were well-designed. The weaknesses of the study consisted of the limited number of stations that the trainees participated in during the clinical performance evaluation and possibly the high-quality MCQ items; however, the results still provide valuable information that can be used to improve the design of the training program such as arranging more core competencies in the 1-year program.

Source of support

Peng-Wei Hsu received funding from this study from Chang Gung Memorial Hospital (grant CDRPG3B0021).

Conflicts of interest

None declared.

Acknowledgments

We would like to thank Liang-Hsiu Ou, MD; Yu-Chih Chang, MD; Yu-Che Chang, MD; Chien-Ta Huang, MD; Chang-Chi Cheng, MD; and Yi-Hao Lin, MD; for their assistance in the development and preparation of test items and the correction of topics. We also thank Ching-Ping Shih and Yi-Chun Chen for their support in human resources and administrative affairs, and Po-Jen Chen, for help with data and statistical analysis.We received funding for this study from Chang Gung Memorial Hospital (grant CDRPG3B0021).

REFERENCES

[1] Kuhlmann TP, Fang WL, Fan Y. Physicians' views on how specialty-specific the first year of residency should be. Acad Med 1991;66:237–9.
[2] Cohen M. The GP preregistration house officer: the potential learning experience of primary care. Hosp Med 1998;59:502–4.
[3] Cassie JM, Armbruster JS, Bowmer MI, Leach DC. Accreditation of postgraduate medical education in the United States and Canada: a comparison of two systems. Med Educ 1999;33:493–8.
[4] Accreditation Council for Graduate Medical Education and American Board of Medical Specialties: Toolbox of assessment methods. ACGME/ABMS joint initiative attachment, Ver. 1.1; September, 2000.

[5] De Champlain AF, Melnick D, Scoles P, Subhiyah R, Holtzman K, Swanson D, et al. Assessing medical students' clinical sciences knowledge in France: a collaboration between the NBME and a consortium of French medical schools. Acad Med 2003;78:509–17.
[6] Haladyna TM, Downing SM. Validity of a taxonomy of multiple-choice item-writing rules. Appl Meas Educ 1989;2:52–78.
[7] Swanson DB, Case SM. Assessment in basic science instruction: directions for practice and research. Adv Health Sci Educ Theory Pract 1997;2:71–84.
[8] Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew RH. The quality of in-house medical school examinations. Acad Med 2002;77:156–61.
[9] Fred P. Validity of the discrimination index as a measure of item quality. J Educ Meas 1973;10:227–31.
[10] Sim SM, Rasiah RI. Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. Ann Acad Med Singap 2006;35:67–71.
[11] Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). Med Educ 1979;13:41–54.
[12] Cleland JA, Abe K, Rethans JJ. The use of simulated patients in medical education: AMEE Guide No 42. Med Teach 2009;31:477–86.
[13] Hodges B, Regehr G, Hanson M, McNaughton N. Validation of an objective structured clinical examination in psychiatry. Acad Med 1998;73:910–2.
[14] Carr S. The Foundation Programme assessment tools: an opportunity to enhance feedback to trainees? Postgrad Med J 2006;82:576–9.