Validity, Reliability and Feasibility of the Observed Long Case Examination as a part of National License Examination for medical undergraduates in Thailand.

Panwara Paritakul MD. MMedSci MedEd[1*]
Wichean Manolertthewan MD.[1]
Piyanant Chonmaitree MD.[2]
Monton Wongwandee MD.[2]

[1] Department of Obstetrics and Gynecology, Faculty of Medicine, Srinakharinwirot University, Bangkok, Thailand
[2] Department of Internal Medicine, Faculty of Medicine, Srinakharinwirot University, Bangkok, Thailand
* Corresponding author, email address: Panwara Paritakul. Department of Obstetrics and Gynecology, Faculty of Medicine, Srinakharinwirot University, Bangkok 10110, Thailand, paritakulp@gmail.com +66 8-3073-6145
Vajira Med J 2015; 59(3): 35-43
http://dx.doi.org/10.14456/vmj.2015.17

Abstract

Since 2007, the observed long case examination has been included in the National License Examination in Thailand. Despite its good face validity and uniqueness in ability to assess the candidate's clinical performance, the long case examination, according to previous literatures, has poor reliability when compared with the objective structured clinical examination (OSCE). Strategies that may improve reliability of long case examination include increasing the number of cases and examiners in each examination and increasing the number of cases but has to be weighed against lowering the feasibility. In summary, for high stake examination, it is reasonable to use the observed long case examination in conjunction with other examination methods such as OSCE to evaluate clinical competency of a candidate.

Keywords: Observed long case, national license examination

ความเที่ยง ความเชื่อมั่น และความเป็นไปได้ของการสอบรายยาวแบบมีผู้สังเกตการณ์เพื่อเป็นส่วนหนึ่งสำหรับการสอบประเมินความรู้ความสามารถในการประกอบวิชาชีพเวชกรรมของประเทศไทย

พรรณวรา ปริตกุล พ.บ. วว. สูติศาสตร์-นรีเวชวิทยา MMedSci MedEd[1*]

วิเชียร มโนเลิศเทวัญ พ.บ. วว. สูติศาสตร์-นรีเวชวิทยา[1]

ปิยนันท์ ชนไมตรี พ.บ. วว. อายุรศาสตร์[2]

มณฑล ว่องวันดี พ.บ. วว. อายุรศาสตร[2]

[1] ภาควิชาสูติศาสตร์-นรีเวชวิทยา คณะแพทยศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ ประเทศไทย

[2] ภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ ประเทศไทย

* ผู้ติดต่อ, อีเมล์: พญ. พรรณวรา ปริตกุล ภาควิชาสูติศาสตร์-นรีเวชวิทยา คณะแพทยศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพ 10110, **paritakulp@gmail.com** โทรศัพท์ 083-0736145

Vajira Med J 2015; 59(3): 35-43

http://dx.doi.org/10.14456/vmj.2015.17

บทคัดย่อ

ตั้งแต่ปีพุทธศักราช 2550 ได้มีการนำการสอบรายยาวเข้ามาเป็นส่วนหนึ่งในการสอบเพื่อรับใบประกอบวิชาชีพเวชกรรมในประเทศไทย ซึ่งการสอบรายยาวนั้น แม้จะมีความเที่ยงตรงที่ปรากฏภายนอก (face validity) ที่สูง และมีลักษณะการสอบที่มาสามารถสังเกตและวัดความสามารถทางคลินิกของนิสิตนักศึกษาแพทย์ได้เป็นอย่างดี แต่จากการทบทวนวรรณกรรม การสอบรายยาวนั้นกลับมีค่าความเชื่อมั่นที่ค่อนข้างต่ำหากเปรียบเทียบกับการประเมินทักษะทางคลินิกด้วย objective structured clinical examination (OSCE) กระบวนการที่อาจเพิ่มค่าความเชื่อมั่นของการสอบรายยาวนั้นได้แก่ การเพิ่มจำนวนอาจารย์ผู้คุมสอบต่อการสอบแต่ละครั้ง และการเพิ่มจำนวนรายผู้ป่วยที่ใช้สอบ แต่กระบวนการเหล่านี้ก็ต้องแลกมากับความยุ่งยากในการสอบ ซึ่งอาจลดความเป็นไปได้ (feasibility) ในการจัดสอบลง โดยสรุปแล้ว การสอบที่มีเดิมพันสูง เช่น การสอบเพื่อรับใบประกอบวิชาชีพ ควรใช้การประเมินร่วมกันหลายด้านทั้งการสอบรายยาวและการสอบ OSCE

Introduction

The Long Case examination is a traditional method of assessing candidate's clinical competence when interacting with real patients. The format of the long case examination may vary. The "traditional long case" requires the candidate to spend unobserved time with a patient, during which the candidate perform history taking and physical examination. The candidate then report back to the examiner and discuss his or her findings, diagnosis and treatment plans with the examiner.[1] The competence of the candidate is then judged based on this discussion. The "observed long case" adds the element of the examiner directly observing the candidate during patient encounter. The ability of a candidate to interact appropriately with the patient is taken into account to pass or fail the exam.[2]

The strength of the long case examination is its authenticity because it involves interacting with real patients under real clinical environment. Another reason supporting its use is the ability to evaluate the candidate at the "show-how" or "does" level of Miller's pyramid[3]. Although this assessment is highly valued and its use as a formative assessment is appreciable, its use as a summative purpose of high-stakes examination is still questionable because of concerns regarding its reliability and validity[2, 4].

In order to obtain a license for medical practice in Thailand, medical students have to pass three parts of national license examination (NLE) held by the Center for Medical Competency Assessment and Accreditation of the Thai Medical Council (CMA Thai), including two observed long case examination which is held by individual medical school or teaching hospital. However, as mentioned above, the reliability and validity of the long case examination itself is arguable. In addition, this long case examination is done at individual medical school, as a result, the robustness of this examination is even more questionable. The aim of this article is to review the conduct of the long case examination in Thailand and discuss the validity, reliability and feasibility of this observed long case examination with regards to the relevant literatures.

The National License Examinations for medical graduates in Thailand

Starting in the year 2007, to qualify for medical practice in Thailand, medical graduates have to pass the 3 parts of NLE[5], which are:

1. NLE part 1: A paper-based 300 multiple choice questions examination covering basic medical science.
2. NLE part 2: A paper-based 300 multiple choice questions covering clinical science.
3. NLE part 3: A 20-station OSCE covering five skills domain (History taking skill, Physical examination skill, Procedural skill, Communication skill, and Interpretation of investigations), and long case examination,

For the long case examination, two cases are required. One from Internal medicine and another from Obstetrics and gynecology, Surgery or Pediatrics. The first two parts of NLE is held nation-wide twice a year and is regulated by the CMA Thai. However, for the long case examination, the CMA Thai only provide instructions, but details such as when and how to perform the examination depend on medical schools.

The process of conducting an observed long case examination

During clinical placements, the 6th year medical students apply for a long case examination with the discipline that they want to be evaluated. The Internal medicine long case exam is mandatory, but the student can choose another case from one of the major disciplines that they prefer (OB-GYN, Surgery, or Pediatrics). The examiners will be a randomly assigned clinical instructor in that discipline. The CMA Thai does not provide regulation regarding the number of the examiners. Hence, some medical schools with limited resources of faculty members may employ only one examiner for each long case examination. The standard procedures for a long case examination, as suggested by the CMA Thai are of the following:

The long case examination should take approximately 60 minutes in total.

Part I: Patient encounter under direct observation: 30 minutes. This time is provided for history taking, physical examination and Initial patient counselling and response to patient's questions.

Part II: Student preparation: 5 minute. This time is provided for summarisation of data and preparation for case discussion with the examiner.

Part III: Case presentation and discussion: 20 minutes. During this time, the examiner assesses the candidate with regards to the ability to summarise pertinent data and problem list, discuss each problem using SOAP model (subjective data, objective data, assessment, plan), give differential diagnosis, diagnostic and treatment plans and describe appropriate patient education

Part IV: Evaluation and feedback: 5 minutes. During this time, the examiner assesses the candidate performance and provides constructive feedback.

The mark will be given on the standard scoring sheet provided by the CMA Thai. The scoring scheme is presented as a table with rigid scoring system (Table 1). The examiner can only check the boxes with the most agreeable scores given.

The passing score is set at 60% by the CMA Thai. If candidates fail, they can apply for re-examination during their clinical placement on other disciplines. This process can be repeated until the candidate passes.

The validity of the observed long case examination

The observed long case examination possesses a good face validity because the process resembles daily work of a clinician. The candidate needs to obtain relevant information from a patient, structures problems, synthesises the findings and formulates a management plan[1]. Because performance of candidates are assessed with real patients instead of simulated patients, the experience is more authentic as it provides a real-time, actual patient problems and a holistic approach[2].

Looking at the concurrent validity, another method that is commonly used to assess candidates' clinical practice at "show-how" level is the Objective Structured Clinical Examination (OSCE). The OSCE was introduced around 1970's and was well accepted in the field of medical education. There are attempts to identify the relationship between the long case and the OSCE[6-8]. Overall, the correlation between the OSCE and the long case result is poor. In the most recent study by Kamarudin, the result showed that there was a positive weak correlation between the total long case mark and the total OSCE mark (r=0.168, p= 0.039). This may result from the fact that the long case examinations in this study are traditionally conducted with unobserved history taking and physical examination while the OSCE type examination requires the candidate to be thoroughly observed. This poses a question whether adding observation of history taking and physical examination part would improve the positive correlation between the long case and the OSCE.[7] Unfortunately, to date, there is still no study to investigate the correlation between the observed long case and the OSCE scores. Moreover, if we take a view from another angle, is it possible that the OSCE and long case examination assess the candidate in a different domain, so the results of these two exams are poorly correlated. In conclusion, with the lack of evidence to support concurrent validity of the long case, the long case should be used in conjunction with other examination methods in a high stake examination.

The reliability of the observed long case examination

Over the past 10 years, the OSCE has replaced most of long case examination in high stake examination[9] because of the ambiguity in the reproducibility of results obtained from the long cases. The low reliability of the traditional unobserved long cases has been mentioned widely[10-12]. Adding direct observation to the history

Table 1:

A standard scoring sheet for long case examination provided by the CMA Thai

Evaluated subject	Score weight	Very good (Accurate data and overall performance > 80%)	Good (Accurate data and overall performance 60-80%)	Improvement required (Accurate data and overall performance < 60%)
1. Data gathering				
1.1 History taking	20	☐ 20 Complete history, all important data that is relevant to patient's problem was gained. The interview is systematically done and includes Personal History, Family History and Review of Systems	☐ 16 ☐ 12 Acceptable. Relevant data was gained. The interview is systematically done.	☐ 8 ☐ 4 Important data is missed. The data gained is not relevant to the patient's problem.
1.2 Physical examination	20	☐ 20 Correct examination techniques. The exam is systematically done. Focus on the areas that are related to the patient's problem. The examination result is mostly accurate.	☐ 16 ☐ 12 Correct examination techniques. The exam is systematically done. Focus on the areas that are related to the patient' s problem. The examination result is acceptably accurate	☐ 8 ☐ 4 Most examination techniques are incorrect. The exam is not done systematically. Focus is not given to the area related to the patient's problem. The examination result is mostly wrong.
2. Case presentation, analysis and clinical reasoning	20	☐ 20 The presentation is done systematically. Very precise and get to the point. All pertinent data is summarised. The differential diagnosis is appropriate for demonstration of good clinical reasoning.	☐ 16 ☐ 12 The presentation is done systematically. Precise and get to the point. Most pertinent data is summarised. The differential diagnosis is appropriate for demonstration of adequate clinical reasoning.	☐ 8 ☐ 4 The presentation is not systematically done. The patient's problem is not analysed. The differential diagnosis is irrelevant with demonstration of poor clinical reasoning.
3. Patient problem-solving including diagnostic plan, therapeutic plan and patient education.	15	☐ 15 Demonstrate good ability to give reasonable investigational plan and initiate appropriate treatment plan for the patient. Investigational results are, most of the time, interpreted correctly. The candidate uses holistic approach and comprehensively describe prevention, rehabilitation and health promotion for the disease.	☐ 12 ☐ 9 Demonstrate acceptable ability to give reasonable investigational plan and initiate appropriate treatment plan. . Investigational results are, in average, interpreted correctly. The candidate can adequately describe prevention, rehabilitation and health promotion for the disease.	☐ 6 ☐ 3 The investigational plan and treatment plan is not appropriate. The investigational results are incorrectly interpreted. The candidate demonstrates little understanding of prevention, rehabilitation and health promotion for the disease.

Vajira Medical Journal
Vol. 59 No. 3 July – September 2015

4. Communication with the patient	15	☐ 15	The candidate introduces himself and asks for patient's consent. The purpose of this interview and examination was addressed. Appropriate types of questions are used. Demonstrate two ways communication and answer patient's question correctly most of the time. Use both verbal and nonverbal language.	☐ 12 ☐ 9	The candidate introduces himself and asks for patient's consent, but the purpose of this interview and examination was not addressed. Demonstrate two ways communication and answer patient's question correctly on average.	☐ 6 ☐ 3	The candidate does not introduce himself nor ask for patient's consent. The purpose of this interview and examination was not addressed. Use only closed- ended questions. Demonstrate one-way communication.
5. Professional attitude and etiquette	10	☐ 10	Respect patient's right. Demonstrate great caring attitude towards patient and patient's relatives. The candidate dress appropriately and exhibit good bedside manner. Demonstrate a good temperamental control and always respond to the patient appropriately.	☐ 8 ☐ 6	Respect patient's right. Demonstrate a caring attitude towards patient and patient's relatives. The candidate dress appropriately and exhibit average bedside manner. Demonstrate an acceptable temperamental control and respond to the patient appropriately.	☐ 4 ☐ 2	Shows little respect to patient's right. Demonstrate little caring attitude. The candidate dress inappropriately and exhibit poor bedside manner. Demonstrate less than the acceptable temperamental control and respond to the patient inappropriately.
Overall performance		**Very good** ☐		**Good Average** ☐☐		**Fail** ☐	

taking component has proved to increase reliability. However, there are many other factors that still contribute to its unreliability, such as, variability in a case difficulty and examiner subjectivity. Various modifications to the traditional long cases have been proposed in order to improve the reliability of the test[2]. The proposed strategies are to increase number of examiners and cases, use a structured question grid[13] or replace with Objective Structured Long Examination Record (OSLER)[10]. By far, only increasing the number of examiners and number of cases are supported by evidence.

A study by Wass V. et al[14] reported a high reliability of the long case examination at 0.84, a considerably high figure in contrast to other previous studies. They compared the reliability between 10 long cases examinations and a 30 stations OSCE examination. The reliability of the long case was 0.84 and of the OSCE was 0.72. They conclude that long case examinations are, in terms of reliability, no worse and no better than OSCEs in assessing clinical competence[14]. However, it is important to note that their results were based on multiple long case examinations (10 cases, spending 20 minute each) which is not practical for most medical schools to conduct such extensive long case examinations.

Even though the long case as a part of Thai NLE is the "observed" type, which proved to have more reliability, there are still other factors that may influence the robustness of this examination method.

Inter-rater reliability

In order to avoid the long case examination being just "luck of the draw", there should be at least two examiners for the examination to weight their judgement. Long case examinations were used at of board certifying level in Thailand for The Royal College of Physicians of Thailand (RCPT) and The Royal College of Obstetricians and Gynecologists (RTCOG). Both colleges rely on 2 trained examiners to evaluate a candidate. Recently, the RCPT reports the high inter-rater reliability with intraclass correlation between

examiners of 0.71-0.83[15].

However, the CMA Thai does not give formal instruction in the number of examiners that should be involved in a long case examination for NLE examination. This leads to some medical schools employ only one faculty member to each long case examination to minimize staffs' burden. Undeniably, a long case examination with only one examiner would trigger a question of its fairness. A student that could "Pass" with one examiner might be regarded as "Fail" to the other more stringent examiner. Especially, when there is no adequate training to standardize all examiners on how to mark a candidate. The CMA Thai does hold several workshops or seminars to provide faculty members opportunities to meet up and share their experience with long case examinations at their own medical schools in order to improve the process. However, these workshops/seminars are optional. Moreover, those who do not involve in medical education rarely attend. Another interesting fact is that newly recruited faculty members who never attend any kind of training from CMA Thai are still eligible to be examiners at their own medical schools. Although a question regarding the reliability of using only one examiner is imminent, this issue is rarely raised by the candidates. Because as mentioned before, a candidate can attempt the long case examination more than once during their clinical placement. If they fail with one examiner, they can simply apply for another exam in other disciplines with a different examiner. This solution, if viewed form another angle, is simply adding case numbers to improve reliability of the long case examination. However, the summative purpose of this long case examination may get mixed up with formative purpose if students' perception towards the examination is not as attentive as it should be.

Test-Retest reliability (Case variability)

Previous studies have shown that increasing number of cases for long case examination will increase reliability of the test. If 2 examiners are involved, assessing with 2 and 3 and 4 cases gives the dependability coefficient of 0.6, 0.69 and 0.75 respectively.[4] However, there is still question regarding the candidates' ability to select their preferred discipline. If a candidate pass two long cases examination in Internal medicine and surgery discipline, does it means he is competent as a doctor? Inability to sample across the broad range of disease from all disciplines in a long case examination is another weakness of this examination. The CMA Thai requires that a candidate must pass one long case examination in Internal Medicine discipline, but has a freedom to choose another case from Pediatrics, Surgery or Obstetrics and Gynecology. This model does not replicate the real medical practice as one cannot select patients with diseases that he is familiar with in real life.

The feasibility of the observed long case examination

The feasibility of the observed long case examination for medical undergraduates in Thailand has been proved to be fair, as the examination process has been continued for many years. However, as previously discussed, the examination still does not meet the optimal standard as some medical school assign only one examiner for each examination to trade off with a more agreeable workload of their faculty members. If the CMA Thai calls for a regulation that two examiners must be allocated for each long case examination, there is likelihood to be objected by faculty members in some medical schools.

Number of cases is also another important factor to improve the reliability of the long case. This requires many efforts, money and time to be spent. With expanding number of 6th year medical students each year, finding adequate number patients who are suitable for long case examination and are willing to engage in such a time-consuming activity is getting more difficult. With limited number of cases, chances are the same patient being called over and over to participate in an examination. Such action would decrease the variation of case selection, and

students might be able to assume in advance which patients or disease they are going to have for their examination. One way to recruit more patients is to provide the appropriate incentive for the patients. At our institution, we offer an inconvenience fee of 300 THB (6 GBP) for each patient who participates in the examination. A consideration to raise the inconvenience fee may make it easier to gain patient consent. However, this must be weighed against the money to be invested.

Conclusion

Despite the uncertainty in its reliability, the observed long case examination still holds its uniqueness in ability to assess the candidate's clinical performance and critical thinking process. Therefore, it is reasonable that the CMA Thai use the observed long case examination in adjunct with other examination methods (OSCE) to assess clinical competence of a candidate on a high stake licensing examination. However, there are still some possible improvements to achieve a higher standard in the long case examination. Firstly, as increasing the number of examiners might prove to be problematic, focus should be aimed towards quality assurance of the examiners. All faculty members, especially the newly recruited, should attend at least one formal training from CMA Thai in order to qualify to be an examiner. The formal training may be held once or twice a year and includes lecture or small group discussion regarding a good conduct for an observed long case examination, the expected roles for examiners, the description of each scoring categories, and familiarize the examiners with instruction of the scoring sheet. This will ensure that all examiners involved in the long case examination have the same criteria to assess a candidate. Second, The CMA Thai should provide an open opportunity for faculty members to feedback and report problems of using the standard scoring sheets. Moreover, their comments could be reviewed to revise the scoring sheet periodically. Finally, despite many years of implementing the observed long case examination in the process of NLE, the examination results have never been analysed. Further study to look at the reliability and correlation of the long case examination results with other clinical performance assessment may contribute to a deeper insight of the validity of current practice and possibly have great impact on decision whether to continue this NLE strategy or to implement a new one.

References

1. Norcini JJ. The death of the long case? BMJ. 2002; 324(7334): 408-9.
2. Ponnamperuma GG, Karunathilake IM, McAleer S, Davis MH. The long case and its modifications: a literature review. Med Educ. 2009; 43(10): 936-41.
3. Miller GE. The assessment of clinical skills/competence/performance. Academic medicine : journal of the Association of American Medical Colleges. 1990; 65(9 Suppl): S63-7.
4. Wilkinson TJ, Campbell PJ, Judd SJ. Reliability of the long case. Medical education. 2008; 42(9): 887-93.
5. The Medical Council of Thailand. Practice in Thailand - Licensing examination. 2014 [3 March 2014]; Available from: http://www.tmc.or.th/en_subcommittee-examination.php.
6. Kamarudin MA, Mohamad N, Siraj MNABHH, Yaman MN. The Relationship between Modified Long Case and Objective Structured Clinical Examination (Osce) in Final Professional Examination 2011 Held in UKM Medical Centre. Procedia - Social and Behavioral Sciences. 2012 ; 60: 241-8.
7. Wass V, Jolly B. Does observation add to the validity of the long case? Medical education. 2001; 35(8): 729-34.
8. Probert CS, Cahill DJ, McCann GL, Ben-Shlomo Y. Traditional finals and OSCEs in predicting consultant and self-reported clinical skills of PRHOs: a pilot study. Medical education. 2003; 37(7): 597-602.

9. van der Vleuten C. Validity of final examinations in undergraduate medical training. Bmj. 2000; 321(7270): 1217-9.
10. Gleeson F. AMEE Medical Education Guide No 9. Assessment of clinical competence using the Objective Structured Long Examination Record (OSLER). Med Teach. 1997; 19(1): 7-14.
11. Norcini J. The validity of long cases. Medical education. 2001; 35(8): 720-1.
12. Paul VK. Assessment of clinical competence of undergraduate medical students. Indian journal of pediatrics. 1994; 61(2): 145-51.
13. Olson LG, Coughlan J, Rolfe I, Hensley MJ. The effect of a Structured Question Grid on the validity and perceived fairness of a medical long case assessment. Medical education. 2000; 34(1): 46-52.
14. Wass V, Jones R, Van der Vleuten C. Standardized or real patients to test clinical competence? The long case revisited. Med Educ. 2001; 35(4): 321-5.
15. Chierakul N, Danchaivijitr S, Kontee P, Naruman C. Reliability and validity of long case and short case in internal medicine board certification examination. Journal of the Medical Association of Thailand = Chotmaihet thangphaet. 2010; 93 (4): 424-8.